



FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE

Ústav anglického jazyka a didaktiky

Department of the English Language and ELT Methodology

Název práce: Opakující se slovní spojení v akademických přednáškách

Title of thesis: Lexical Bundles in Academic Lectures

BAKALÁŘSKÁ PRÁCE / BA THESIS

Vedoucí bakalářské práce (supervised by):	Zpracoval (written by):
PhDr. Markéta Malá, Ph.D.	Marek Leško

Praha, srpen 2009 / Prague, August 2009

„Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a že jsem uvedl všechny použité prameny a literaturu. Souhlasím se zapůjčením bakalářské práce ke studijním účelům.“

“I declare that the following BA thesis is my own work for which I used only the sources and literature mentioned. I have no objections to the BA thesis being borrowed and used for study purposes.”

Místo, datum a podpis (place, date and signature):

Acknowledgment

I would like to thank my supervisor, PhDr. Markéta Malá, Ph.D., for her thorough and sound advice as well as a patient and supportive attitude even under time constraints.

Contents

1. Introduction.....	5
2. The definition of lexical bundles	5
3. General characteristics of lexical bundles	7
3.1. Lexical bundles vs. idioms	7
3.2. Structural completeness.....	7
3.3. Combination, sequences and overlaps.....	8
3.4. Lexical bundles – the summary of their use from a quantitative point of view	9
4. The structural characteristics of lexical bundles.....	10
4.1. The difference between “conversational” and “academic” bundles.....	11
4.2. The structural patterns of lexical bundles in conversation	12
4.3. The structural patterns of lexical bundles in academic prose.....	17
5. The functions of lexical bundles.....	21
5.1. Lexical bundles as stance expressions.....	22
5.1.1. Epistemic stance bundles	22
5.1.2. Attitudinal or modality stance bundles.....	23
5.2. Lexical bundles as discourse organizing devices	23
5.2.1. Topic introduction/focus bundles.....	24
5.2.2. Topic elaboration/clarification bundles.....	24
5.2.3. Identification/focus bundles	24
5.3. Referential bundles	25
5.3.1. Imprecision bundles	25
5.3.2. Bundles specifying attributes	25
5.3.3. Bundles referring to time, place or text.....	26

5.4.	Distribution of lexical bundles in different types of discourse.....	26
6.	The cohesive role of lexical bundles.....	27
7.	Research of the lexical bundles in academic lectures.....	28
7.1.	Research scope and methods.....	28
7.2.	Research data: overview of the most frequent bundles.....	30
7.3.	Idiosyncratic bundles.....	33
7.4.	Shorter and longer bundles.....	34
7.5.	Structural categorization.....	34
7.6.	The discourse functions of the lexical bundles in the transcripts.....	35
7.6.1.	Stance bundles.....	37
7.6.2.	Discourse organizers.....	42
7.6.3.	Referential expressions.....	44
7.7.	Cohesive role of lexical bundles in academic lectures.....	46
8.	Conclusions.....	49
9.	Summary of the work in Slovak (Zhrnutie práce v slovenskom jazyku)	50

1. Introduction

Lexical bundles, or multi-word sequences, have been studied under several different names, including ‘lexical phrases’, ‘formulas’, ‘routines’, ‘fixed expressions’ or ‘pre-fabricated patterns’ (‘prefabs’) (Biber and Barbieri 2007, 264), and this variation in terminology reflects different approaches to the subject. The term ‘lexical bundle’ appeared first in the Longman Grammar of Spoken and Written English and a series of subsequent studies continued the research within that framework (Biber 2006, 129). The importance of this area of study has been demonstrated on several different levels. As will be shown, a significant portion of language production consists of recurrent expressions, which is an observation valid for both spoken and written texts (Biber 1999, 993 – further referred to as LG). The extent of their use results in practical necessity for language learners to acquire the ability to understand and use these multi-word sequences. The role of lexical bundles in second language acquisition is still disputed, however, because even though the importance of formulaic language for fluency cannot be denied, there is little consensus over the introduction of lexical bundles in teaching curricula (Biber and Barbieri 2007, 284). The present work will attempt to summarize the linguistic phenomenon of lexical bundles in general, analyze their functions in the academic lectures from the corpora of BASE (British Academic Spoken English) and MICASE (Michigan Corpus of Academic Spoken English) and discuss the differences between British and American English in terms of their use.

2. The definition of lexical bundles

As has been noted, there are several possible approaches to the area of multi-word expressions. Some studies focus on expressions that are idiomatic, i.e. their meaning cannot be derived from the individual words (such as *in a nutshell*), while others focus on the perceptual salience of the expressions (e.g. *you’re never going to believe this*). (Biber and Barbieri 2007, 265) Lexical bundles, however, are a specific category of multi-word expressions, because they are not defined by their idiomatic nature or perceptual salience but rather statistical occurrence. Thus, Longman Grammar defines lexical bundles as “extended collocations: bundles of words that show a statistical tendency to co-occur” (LG, 989). Because it might be useful to take into

consideration the differences between various fields and modes of discourse – an expression that functions as a lexical bundle in one register might not occur at all in another – lexical bundles could be also defined as “multi-word sequences that occur most commonly in a given register” (Biber and Barbieri 2007, 264). It should be noted that “most commonly” is a vague term that is usually clarified as a statistical boundary for the purposes of a particular study: Longman Grammar of Spoken and Written English, for instance, sets the limit at 10 occurrences per million words for four-word bundles, whereas recent research usually chooses a more conservative frequency limit of 40 times per million words (Biber and Barbieri 2007, 267). Even though two-word expressions that occur often enough should fall under the definition of a lexical bundle, most studies consider only bundles that consist of three and more words, because shorter bundles are often incorporated in the longer ones, the high frequency of many two-word expressions (*it is, even though, as well, etc.*) would make the research less manageable (LG, 990) and finally, longer bundles are more interesting linguistically.

The study of lexical bundles can thus be understood as similar and related to the study of collocations, i.e. lexico-grammatical associations between words, with two important differences. First, whereas collocations are usually studied as relations between two words or a word and the lexical units associated with it, lexical bundles are longer structures (Longman Grammar of Spoken and Written English, for example, analyzes lexical bundles as long as six words (LG, 1013)). Second, as it clearly follows from the definition, lexical bundles are identified by statistical means only and the structural and semantic aspects (e.g., whether an expression functions as a complete and meaningful unit) are disregarded. Accordingly, while the expression *I don't think she* would not probably be discussed in textbooks and articles dealing with collocations because of its low perceptual salience and incompleteness (Biber and Barbieri 2007, 283 – 284), it is a lexical bundle that frequently appears in conversation (LG, 1002). Finally, it should be noted that the statistical methods employed in identifying lexical bundles exclude all discontinuous recurrent expressions such as *not only... but also* or *the [comparative adjective] ... the [comparative adjective]*, which might be considered a disadvantage of this approach (Nesi and Basturkmen 2006, 285).

3. General characteristics of lexical bundles

Before considering the most frequent grammatical structures of lexical bundles and their role in spoken and written English, a few general points should be made about the nature of these multi-word expressions.

3.1. Lexical bundles vs. idioms

It has already been suggested that the respective studies of idiomatic expressions and lexical bundles diverge because of the difference of focus. In fact, most idioms do not occur frequently enough to qualify for lexical bundles. The idiomatic expression to *kick the bucket*, for instance, does not appear in the Michigan Corpus of Academic Spoken English (MICASE) in any form (*kick, kicks, kicking, kicked*), whereas the expression *as you can see* was found 37 times in approximately 1.8 million words. While it is true that such an informal idiom is not expected to be used in academic lectures, research by Biber et al. has shown that even common idioms, such as *a piece of cake* and *a slap in the face*, are statistically quite rare, occurring up to 5 times per million words in fiction and even less in other registers (LG, 1025). However, several lexical bundles are idiomatic, such as *the bottom line*, *the big picture* and *on the other hand* (Biber and Barbieri 2007, 270). Finally, lexical bundles are usually not perceptually salient (Biber and Barbieri 2007, 270), which means that they are not likely to appear as a prominent element in discourse. For instance, in the question “*do you want to go dancing?*” consisting of a common lexical bundle *do you want (to)* and the complement *to go dancing*, the focus is likely to be on the complement in terms of stress and interpretation.

3.2. Structural completeness

Structural completeness is another aspect in which idioms and lexical bundles differ. Whereas idiomatic expressions tend to be structurally complete and they can usually be replaced by an alternative lexical expression (*to kick the bucket* = *to die*), most lexical bundles “extend across structural units” (LG, 991), which is one of the reasons why they are not recognized as fixed lexical expressions. This might also explain why they eluded attention (Biber 2006, 135)

until the onset of large-scale corpus studies. For instance, lexical bundles found in conversation often consist of a pronominal subject, a verb phrase and the initial part of a complement clause (*I don't know why* or *I thought that was* (LG, 996 – 997)). “Academic” lexical bundles, on the other hand, often contain a fragment of a noun phrase or a prepositional phrase (*as a result of, the nature of the* (LG, 996 – 997)). Again, this is not a fixed rule and structurally complete lexical bundles can be found both in conversation (*I don't like it* (LG, 999)) and academic discourse (*in the present stud* (LG, 1019)). However, only 15% of lexical bundles in conversation and 5% of those found in academic prose can be regarded as structurally complete (LG, 995). The structural characteristics of lexical bundles in conversation and academic discourse will be discussed in more detail in another section of the study.

3.3. Combination, sequences and overlaps

The studies on the subject of lexical bundles mostly deal with three- to six-word expressions. On the one hand, two-word bundles are extremely common and it is perhaps misleading to speak of a bundle in the case of short expressions such as *it is* or *I think*. On the other hand, those consisting of seven and more words are too rare to be bundles according to the provided definition. In conversation and academic prose, there are almost ten times as many three-word bundles as four-word bundles and the same ratio is valid for the four- and five-word expressions (LG, 993). Longer lexical bundles are frequently formed as an extension or combination of shorter bundles. Longman Grammar gives an example of this process: three-word recurrent expressions *do you want, you want me, want me to, me to do* can be extended to *do you want me, you want me to, want me to do*. These can then be combined into *do you want me to* and *you want me to do* and ultimately *do you want me to do* (LG, 993). Shorter lexical bundles often appear in a sequence without necessarily forming a recurrent unit: *I'm not sure + if they're + they're going* -> *I'm not sure if they're going* (LG, 999). Overlapping sequences, such as the following example, are particularly common in academic prose: [₁ *but* [₂ *it* [₃ *is*₁] *clear*₂]] [₄ *that*₃] [₅ *there are*₄]] [₅ *at least two*₅], which consists of five lexical bundles, namely *but it is, it is clear, is clear that, that there are, at least two* (LG, 999).

3.4. Lexical bundles – the summary of their use from a quantitative point of view

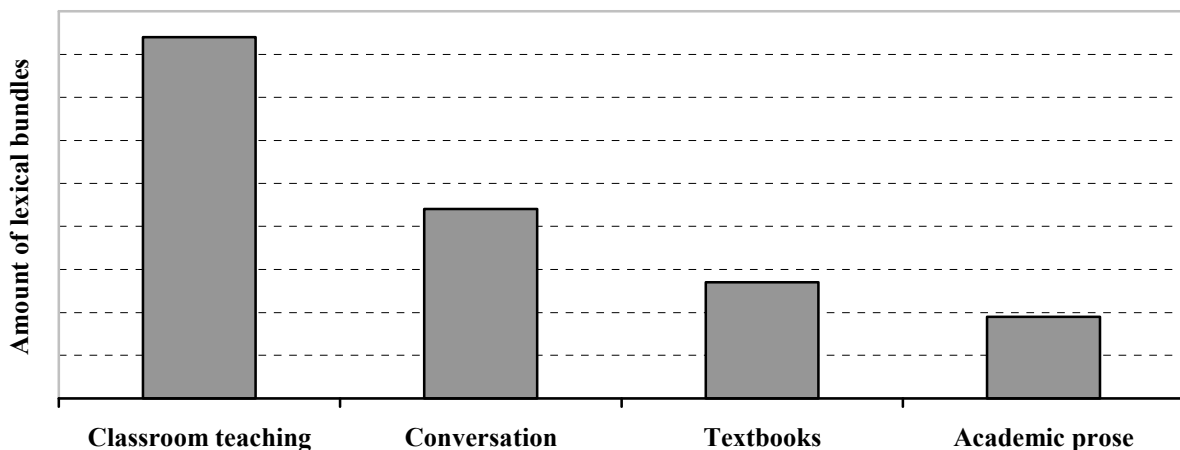
One of the aims of the present study is to analyze the use of lexical bundles in academic lectures from a quantitative point of view. It might be useful, however, to review the findings of previous studies to illustrate the extent of their use and establish the basis for later comparison. In terms of the ratio between lexical bundles and non-recurrent elements, Biber et al. discovered that 25% of conversation consists of three-word bundles, 3% of four-word bundles, 70% being non-recurrent expressions. The remaining 2% are phrasal/prepositional verbs that Biber's study deals with elsewhere as a type of multi-word expressions. In this respect, academic prose differs only slightly: three-word bundles make up 18%, four-word bundles 2%, and non-recurrent expressions 79% of the discourse (LG, 993 – 994). (See Figure 3.4.1. for a more accessible summary of these results.) It should be noted at this point that if the contracted forms are considered lexical bundles, e.g. *I don't* is interpreted as a contracted form of a three-word bundle *I do not*), according to Biber et al., almost 45% of all words in conversation occur in a lexical bundle (LG, 996).

A study by Hyland shows that the use of lexical bundles varies according to the field and disciplines such as biology, electrical engineering, applied linguistics and business studies tend to employ specific lexical bundles, a significant portion of which does not occur in other disciplines (Hyland 2008, 11). Concerning the use of lexical bundles in academic spoken discourse, a study by Biber from 2006 shows that classroom teaching employs almost twice as many bundles as conversation and about four times as many as academic prose (see Figure 3.4.2.) (Biber 2006, 147). In terms of the structural characteristics, a paper by Hilary Nesi and Helen Basturkmen concludes that classroom teaching employs both “oral” and “literate” elements (Nesi and Basturkmen 2006, 290). As the present study focuses on the use of lexical bundles in British and American academic lectures, the role of lexical bundles in the academic spoken discourse will be discussed separately and the data provided here serve only as an illustration and a point of reference.

Figure 3.4.1. Distribution of recurrent and non-recurrent expression in conversation and academic prose (LG, 993 – 994)

	Conversation	Academic prose
Three-word bundles	25%	18%
Four-word bundles	3%	2%
Phrasal/prepositional verbs	2%	1%
Non-recurrent expression	70%	79%

Figure 3.4.2. Illustration of the proportional distribution of lexical bundles in different types of discourse (Biber 2006, 137)



4. The structural characteristics of lexical bundles

Lexical bundles can be categorized on the basis of grammatical units of which they consist. Because these multi-word expressions rarely represent structurally complete units, as has been already noted, the types of clauses and clause fragments that occur in lexical bundles often play a role in this division. This categorization might be seen as performing merely an enumerative function at first; however, as lexical bundles found in conversation and academic prose, respectively, tend to use dramatically different structures according to the research done by Biber et al. (LG, 990 – 1024), these results will be used later in the analysis of lexical bundles in academic lectures, which combine elements of both conversation and academic discourse (Biber and Barbieri 2007, 265).

4.1. The difference between “conversational” and “academic” bundles

Longman Grammar distinguishes between structural patterns that are more widely used in conversation and those used in academic prose. Because the present study focuses on academic lectures from different corpora, it will not consider the detailed statistical results of Biber et al. but provide a brief summary of each structural type instead. The findings of Longman Grammar show that the division is quite neat: only 2% of lexical bundles found in academic prose employ the patterns typical for conversation and only 8% of conversational bundles use the structural type of the academic bundles. However, 11% of bundles found in conversation and 15% of bundles found in academic prose belong to structural patterns used in both registers (LG, 996).

Generally speaking, 90% of lexical bundles found in conversation are parts of declarative clauses or questions, whereas 60% of lexical bundles in academic prose form parts of noun phrases or prepositional phrases (LG, 995). While structurally incomplete lexical bundles found in conversation end with a verb or a function word with almost equal probability, an overwhelming majority of the lexical bundles in academic prose have a function word at the end. The most common of these function words are articles, prepositions and complementizers (LG, 995). It should be also noted that the lexical bundles in conversation often serve as utterance launchers (*I don't think, that's a good idea, are you going to, etc.*), which is rather uncharacteristic of the “academic” bundles (*are shown to be, by the fact that etc.*) (LG, 1018, 1023).

The following categorization relies on the findings of Biber et al. as presented in the Longman Grammar of Spoken and Written English and examples are always quoted from this source unless indicated otherwise. Because Biber et al. focused on four-word bundles, the majority of examples will be of this length. The minimum frequency set by Biber et al. was ten per million words for four-word bundles and five per million words for five- and six-word bundles.

4.2. The structural patterns of lexical bundles in conversation

Longman Grammar of Spoken and Written English recognizes 14 different structural types of lexical bundles that are used in conversation.

Type 1: A personal pronoun + a lexical verb phrase (+ a fragment of a complement phrase)

This type is by far the most prevalent lexical bundle in conversation and it often includes a fragment of a following phrase. The examples of this type include *I don't know what, I think it was, I said to him, I was going to; I don't know what it, I would like to see; you want a cup of tea*. Most of these bundles are used to begin an utterance (Longman Grammar uses the term “utterance launcher” (LG, 1003) to describe the function of such clauses) and the main clause verb in these sequences often expresses a personal stance. Moreover, some bundles in this category, such as *you know what I mean*, are often associated with personal speech habits. (LG, 1002 – 1005).

Type 2: A noun or a pronoun phrase + be +

A number of different kinds of expressions are included in this type: bundles with *it* (*it's going to be, it's not too bad*), bundles with *that* (*that's what I mean, that's a good idea*), *there* (*there's a lot of, there's going to be*), personal pronoun (*I'm going to be, I was in the*) and finally bundles with another noun phrase (*but the thing is, some of them are*). The bundles that begin an existential-locative predicate with *there* are included in this category, because *there* is considered a vicarious subject in this kind of sentences. The demonstrative pronoun *that* is often followed by a clause starting with *what, who, whom, why* or *how* that refers to a previous utterance (*that's what I said*) (LG, 1005 – 1006).

Type 3: A verb phrase with an active verb

This type of a lexical bundle includes expressions with *have* (*have a look at, let's have a look*), *go* (*go to the toilet, have to go and*), *get* or *got* (*get on with it, get rid of*), *put* (*put the kettle on, put them in the*), *see* (*see if I can, see if you can*), *want* (*want a cup of, want me to do*), other modal or semi-modal expressions (*going to be a, have to pay for, used to be a, used to have a*) and other verb phrase expressions (*thank you very much, hang on a minute*). This category is special in that many of its lexical bundles are – much unlike the rest – often idiomatic¹, such as *hang on a minute* or *haven't got a clue* (LG, 1006 – 1007).

Type 4: Yes-no question fragments

The lexical bundles in this type all begin with an auxiliary or a modal verb followed by a pronominal subject; this structure suggests that these expressions form parts of yes/no questions. Bundles in this group typically use the verbs *have* (*can I have a, did you have a*), *got* (*have you got a, have you got any*), *know* (*do you know what, do you know that*), *want* (*do you want to, do you want me*), and *talk* or *tell* (*are you talking about, did I tell you*). Some bundles in this group also use modal or semi-modal verbs (*are you going to, do you have to*), or verbs that do not belong to any of the above-mentioned category (*would you like to, did you see that*). Expressions such as *do you know what I mean*, which has been discussed with regard to combining bundles into longer sequences, and *do you want a cup of* represent 6-word bundles from this category (LG 1007 – 1008).

¹ For the purposes of this study, *idiomaticity* of a phrase refers to the extent to which the meaning of the whole differs from the denotation of its individual components in sum – in other words, the more difficult it is to derive the meaning of an idiom from the individual words, the more idiomatic the expression is. On one end of the spectrum stand phrases such as the pair *on the one hand – on the other hand*, the meaning of which an average language learner might easily guess, and then there are idioms like *to kick the bucket*, which are extremely remote from their literal denotation. The origin of the above-mentioned example, *to kick the bucket*, is obscure and several contradictory explanations, usually involving suicide, pig-slaughter or execution by vigilantes can be found on the internet.

(see: OED online: bucket - <http://dictionary.oed.com/cgi/entry/50028660>,
<http://www.phrases.org.uk/meanings/218800.html>,
<http://primatejournal.blogspot.com/2006/10/origin-of-kick-bucket.html>)

Type 5: Wh-question fragments

This type of lexical bundles can be categorized according to the *wh*-question word that is used (*what*, *where*, *how*), the questions with *what* being the most common. The *wh*-question word is usually followed by a modal or auxiliary verb and a subject pronoun. Among the verbs that are frequently employed with *what* are *do* (*what are you doing*, *what have you done*), *say* (*what did you say*, *what do you say*) or other common verbs (*what do you mean*, *what's the matter with*). Lexical bundles with *where* and *how* include *where are you going* and *how do you know*. Interestingly, Longman Grammar does not list *wh*-questions starting with *why* or *who*, even though some of the common exclamatory idioms might have been expected to appear among lexical bundles (*why on earth*, *who the hell*) (LG, 1008 – 1009).

Type 6: Lexical bundles with wh-clause fragments

The difference between this type and type 5 is that lexical bundles in this category are not a fragment of a direct question but a different kind of *wh*-clause. As such, they can be divided into two groups: verb phrases followed by a *wh*-clause (usually interrogative content clauses) and bundles that start a *wh*-clause (usually nominal content or adverbial clauses). The verbs *know* (*don't know what it*, *know where it is*) and *see* (*see what I mean*, *see what you mean*) are frequently employed in the first subtype. The second subtype can be illustrated by examples such as *what I'm going to*, *what I said to* or the six-word expression *what I was going to say* (LG, 1009 – 1010).

Type 7: Lexical bundles with to-clause fragments

This type can be further divided into verb phrases followed by a *to*-clause and bundles that begin a *to*-clause. The verbs found in the former subtype include *want* (*don't want to go*, *want to do it*), *be able* (*be able to get*, *might be able to*), and *like* (*would like to go*). The latter subtype uses the verbs *go* (*to go to the*, *to go and see*), *do* (*to do with it*, *to do with the*) and some other verbs and adjectives (*to be able to*, *to get rid of*). In this category, the most frequent verbs in lexical bundles are *want* and *go*, in that order (LG, 1010).

Type 8: A verb + a *that*-clause fragment

It should be noted that despite the name, the complementizer *that* is omitted in all lexical bundles of this type in conversation, as *that* is rarely present in declarative content clauses in spoken discourse. Most of these bundles occur with a small group of very common main clause verbs (*think, say, know*) and a personal pronoun as the subject. Bundles such as *I don't think he, I said I would* are typical examples of this type (LG, 1010 – 1011).

Type 9: Adverbial clause fragments

Two main groups can be identified within this type: lexical bundles that form the beginning of an *if*-clause and bundles starting with a “complex subordinator” (LG, 1011) with the structure *as* + an adverb + *as* + a pronoun. The former group can be illustrated by examples such as *if you want to* or *if you've got a*, while bundles such as *as long as you* and *as far as I* belong to the latter. Finally, there are relatively fixed adverbial expressions, such as *(be)cause I want to*. This type, even though it is not particularly numerous, includes some of the longer bundles, e.g. *if you know what I mean* (LG, 1011).

Type 10: Noun phrase expressions

The bundles of this type comprise noun phrases that include a fragment of an *of*-clause functioning as a postmodifier (*at the end of the, the rest of the*) and other noun phrases or noun phrase fragments (*that sort of thing, or something like that*). Many of these bundles describe a location or specific position (*the back of the, the top of the*), some of the noun phrases in this group are used idiomatically (*end of the day, nothing to do with*), and some are used as hedges (*and things like that, and that sort of thing*). Finally, some of these bundles are used for the sake of emphasis concerning the quantity of something (*quite a lot of, nothing to do with*) (LG, 1012).

Type 11: Prepositional phrase expressions

The lexical bundles of this type begin with a preposition and can be divided into two groups. The first group contains prepositional phrases or prepositional phrase fragments that include a noun followed by *of* (*at the end of*, *for the rest of*), and the second group contains other prepositional phrases and phrase fragments (*for a long time*, *on the other side*). The six-word bundle *at the end of the day* that has already been mentioned as one of the long and idiomatic lexical bundles belongs here as well. The expressions of this type often function as time or place adverbials (*at the same time*, *at the back of*) (LG, 1012 – 1013).

Type 12: Quantifier expressions

There are relatively few lexical bundles of this type and all begin with the quantifier *all*. They are generally rather fixed expressions that are mostly used for emphasis (*all of a sudden*, *all over the place*). However, the expression *and all the rest of it*, which is described as “a coordination tag hedge [used] for vague reference” (LG, 1013), also belongs to this type (LG, 1013).

Type 13: Other expressions

Longman Grammar provides several examples of lexical bundles that do not belong to any of the above-listed categories. Some of these are quantifying expressions that are frequently used in conversation (*two and a half*, *three and a half*), and the rest are bundles formed by reduplication (*no no no no*, *and on and on and on*) that are also referred to as iterative (LG, 1013).

Type 14: Meaningless sound bundles

The last type of conversational lexical bundles covers meaningless expressions such as *da da da da* or *la la la la* that are used in conversation often enough to be listed as recurrent expressions. It should be noted that these expressions are meaningless only in the sense of

lacking particular denotation and they perform a variety of roles in spoken discourse, ranging from expressing agreement (*mm mm mm mm*) to carrying a tune (*da da da da*) (LG, 1014).

4.3. The structural patterns of lexical bundles in academic prose

Even without a detailed analysis, it seems clear that the structural patterns of lexical bundles found in academic prose might differ from those typical of conversation. It appears understandable, for instance, that academic prose does not use meaningless sound bundles (except for linguistic studies on the subject of meaningless sound bundles) and the frequency of personal pronouns, particularly the first and second person, drops rapidly as the formality of the genre increases. The results of the study of Biber et al. are, nonetheless, still surprising as they have demonstrated that the lexical bundles used in conversation overlap very little with those used in academic prose. The study identified 12 different categories, which will be briefly introduced.

Type 1: A noun phrase with an *of*-phrase fragment

This numerous type of a lexical bundle consists of a noun phrase followed by a fragment of an *of*-phrase in the role of a post-modifier. They fulfill various functions, such as physical description, identification or specification of some aspect, such as place, size and amount (*the surface of the, the shape of the*), marking the existence or presence of something (*the presence of the, the existence of a*) or identifying abstract qualities (*the nature of the, the value of the*). Finally, these bundles are used to describe processes or events that last over a period of time (*the development of an, the course of the*) (LG, 1015 – 1016).

Type 2: A noun phrase with other post-modifier fragments

The lexical bundles that consist of a noun phrase but do not fall under the previous type are relatively rare. They can be divided into two groups: noun phrases with a post-nominal clause fragment (i.e. a fragment of a dependent clause modifying the noun that it follows) and noun phrases with a prepositional phrase fragment. The former group is represented by expressions

such as *the way in which* or *the fact that the*, while the latter includes expressions *the relationship between the* and *the same way as*. Many of the lexical bundles of this type describe processes, particularly the expressions *the way(s) in which* and *the extent to which*, and some are used to characterize relations of entities: *the relationship between the* or *the difference between the*. Finally, the booster *the fact that the* also belong to this type (LG, 1016 – 1017).

Type 3: A prepositional phrase with an embedded *of*-phrase fragment

Lexical bundles in the category could be catalogued according to the preposition with which they start. Some of these expressions state abstract, logical relations, particularly those using *as* and *in* (*as a result of*, *in the absence of*), but also *on* (*on the basis of*). On the other hand, if the preposition is *at*, these lexical bundles often mark temporal relations (*at the end of*, *at the time of*). Some of these expressions are used to identify time periods and processes (*in the course of*, *in the development of*) (LG, 1017 – 1018).

Type 4: Other prepositional phrases or phrase fragments

All lexical bundles that consist of a prepositional phrase without an *of*-phrase fragment belong to this category. Some are used to identify a particular location or time period (*in the United States*, *in the nineteenth century*), while others refer to the discourse or some of its parts (*in the present study*, *in the next chapter*). Two particular lexical bundles, *at the same time* and *on the other hand*, deserve special attention as they are extremely common and – unlike most lexical bundles – to an extent idiomatic in meaning. The function of *at the same time* is to contrast two statements that are both considered true or plausible in the context of the discourse, whereas *on the other hand* is used when two incompatible arguments are being considered (LG, 1018 – 1019).

Type 5: An anticipatory *it* + a verb phrase or an adjective phrase

The first subtype, anticipatory *it* followed by a verbs phrase is the less common of the two and includes bundles such as *it can be seen* and *it was found that*. The second subtype, anticipatory *it* followed by adjective, can be illustrated by examples like *it is possible to* and *it is clear that*. Longer bundles, such as *it is interesting to note that* and *it should be noted that the* can be found in this category, too. Most of the adjectives in such bundles are complemented, usually by a *to*-clause, and tend to express epistemic stance (possibility or probability – *it is possible to*) or attitudinal stance (importance or necessity – *it is necessary to*). Unlike the lexical bundles in this type with an adjective phrase, the bundles with a verb phrase or phrase fragment are mostly followed by *that*-clauses and often express the stance that the following information cannot be disputed or denied (*it should be noted that, it can be seen that*) (LG, 1019 – 1020).

Type 6: A passive verb + a prepositional phrase fragment

It has been noted that only few lexical bundles in academic prose are built around verb phrases. Those that belong to this group usually employ passive voice and a subsequent prepositional phrase, such as *are shown in figure, be found in the*. Like in this example, the prepositional phrase tends to refer to a location or a logical relation rather than an agent using a *by*-clause. Besides referring to a part of the written discourse, usually a graph or a table, the lexical bundles of this type “[identify] the basis of some finding or assertion”, e.g. *is based on the (fact)* (LG, 1020 – 1021).

Type 7: The copula *be* + a noun phrase or an adjective phrase

Two main groups can be identified in this type. The first group comprises bundles in which the copular verb is followed by a noun phrase (*is one of the, is the result of*), while the second group includes all those recurrent expressions where an adjective phrase follows the copula (*is due to the, is similar to that*). The difference between this type and type 3 often consists merely in the copular verb; therefore, there are some noticeable overlaps between these categories (cf. *as a result of* – *is a result of*). These lexical bundles often include a subject complement or its

fragment: *is part of the, is one of the*. The bundles containing an adjective phrase express causal relations (*may be due to*) or comparative relations (*is equal to the*) (LG, 1021).

Type 8: (A verb phrase) + *that*-clause fragment

The lexical bundles of this type include a verb phrase followed by a *that*-clause (*should be noted that, has been shown that*) and some of these expressions even contain a noun phrase (*studies have shown that*), but they may be limited to a *that*-clause fragment (*that there is a, that it is the*). These bundles are usually “a part of a larger extraposition structure” (LG, 1022), such as a clause with an anticipatory *it*, which is a structural pattern that has already been discussed. For this reason, overlaps occur between these two categories, too (e.g. *it should be noted* vs. *should be noted that*). On the other hand, the bundles containing only a *that*-clause often appear as declarative clause fragments (*that there is a, that it is not*) (LG, 1022).

Type 9: (A verb or an adjective) + a *to*-clause fragment

This type can be divided into three subgroups: (1) a predicative adjective + a *to*-clause (*are likely to be, is not possible to*); (2) a verb phrase + *to*-clause (*has been shown to, does not seem to*); and finally, (3) a simple *to*-clause (*to be able to, to do with the*). The bundles in the first group are used to express the stance of possibility or ability (*is likely to be, should be able to*). On the other hand, some of the bundles with a verb phrase function as a reference to previous discourse or known information (*was found to be, has been shown to*) (LG, 1022 – 1023).

Type 10: An adverbial clause fragment

According to the findings of Longman Grammar, only four lexical bundles begin with an adverbial clause fragment: *as shown in figure/fig., as we have seen, as we shall see, if there is a*. The function of the bundles starting with *as* can be described as “a deictic reference to other discourse segments” (LG, 1023).

Type 11: A pronoun or a noun phrase + be

Many of the longer bundles fall under this category, such as *this is not to say that, there was no significant difference between, the aim of this study was*. Generally, they can be divided into (1) those starting with *this* (*this is not the, this is not to*) and *there* (*there was no significant, there has been a*). In accordance with intuitive expectations, bundles in academic prose do not employ personal pronouns. *This* at the beginning of a phrase often refers to the preceding discourse (*this is not to say that*), while *there* is often used in phrases focusing on correlation or statistics (*there was no significant difference*) (LG, 1023 – 1024).

Type 12: Other expressions

At the end of the chapter dedicated to lexical bundles, Biber et al. list a few lexical bundles that do not fit into any of the above-mentioned categories: *as well as the, as well as in, than that of the, may or may not, the presence or absence* (LG, 1024).

5. The functions of lexical bundles

Because every categorization or typology is to a certain degree arbitrary and, more importantly, adjusted to the purposes of a particular study, the various functions of lexical bundles can be approached from several different angles. Biber's most recent studies rely on the basic distinction between three primary discourse functions of lexical bundles, namely (1) *stance expressions*, (2) *discourse organizers* and (3) *referential expressions*, all of which can be further divided into several subcategories (Biber and Barbieri 2007, 270). Hyland, on the other hand, makes the basic distinction based to the focus of bundles, which can thus be divided into (1) *research oriented*, (2) *text-oriented* and (3) *participant-oriented* (Hyland 2008, 12). These two systems partly overlap: Hyland's participant-oriented bundles (*it is possible that*) are mostly stance expressions in Biber's terminology, the research-oriented bundles (*in the Hong Kong*) roughly correspond to referential expressions and the text-oriented bundles (*in addition to the*) largely function as discourse organizers (Hyland 2008, 13 – 14). Because Biber's division is more general and suited to a variety of spoken and written genres, whereas Hyland's was used to

compare specific fields of academic written discourse (research articles, Master theses and doctoral dissertations) with regard to the use of lexical bundles, the present study will adopt Biber's typology. The examples are quoted from Biber's *University Language* (2006) unless indicated otherwise.

5.1. Lexical bundles as stance expressions

The stance bundles "provide a frame for the interpretation of the following proposition" (Biber 2006, 135) and as such, they can convey two major types of meaning that modify a portion of the discourse. The epistemic stance bundles express the speaker's or author's attitude towards the probability or possibility of the information, whereas the attitudinal/modality stance bundles convey the speaker's or writer's attitude towards the content of the affected propositions. From the formal perspective, stance expressions can be divided into *personal* and *impersonal*. The former explicitly refer to the author or the speaker, whereas the latter are not directly attributed to the source of the text or utterance (Biber 2006, 135). It should be noted, however, that even impersonal stance expressions can fulfill the role of attitudinal stance markers that convey the subjective attitude of the speaker (*it should be noted* or *it must not be said*, for instance).

5.1.1. Epistemic stance bundles

Research by Biber (2006) indicates that the majority of stance bundles are personal, especially in classroom teaching, and that they mostly express what might be termed negative epistemic modality, i.e. uncertainty, improbability or lack of knowledge (*I don't know what, I don't know if, I don't think so*). In some cases, this kind of bundles combines the function of expressing stance with referential identification, such as *I think it was [the]* and *stuff like that*, the latter being an example of an imprecision bundle that is still referential to a certain degree. In contrast with the personal epistemic stance bundles, the impersonal expressions tend to express positive epistemic modality, i.e. certainty, probability or verified information (*are more likely to, the fact that the*) (Biber 2006, 140).

5.1.2. Attitudinal or modality stance bundles

Stance bundles of this type are usually personal and convey the speaker's attitude towards the contents of the relevant proposition, which usually follows the stance bundles. Biber distinguishes four major subcategories: (1) *desire*, (2) *obligation* and *directive*, (3) *intention* and *prediction*, and (4) *ability*. The first of these, attitudinal stance of desire, include only personal expressions that convey the speaker's wish or ask about the desires of other participants in the discourse (*I don't want to, what I want to do, do you want me to*). The second group, obligation and directive stance bundles, includes expressions such as *you need to know* and *you have to do*. They mostly use the second person pronoun, but they can still be understood as bundles of stance on the speaker's part directing other participants to carry out certain instructions. They can range from very direct (*I want you to*) to rather oblique (*you might want to*). Some of these bundles also fulfill the function of topic introduction (*take a look at*). The intention and prediction includes many overtly personal bundles that convey the speaker's plans or future objectives. They often express joint action (*what we're going to*) or, in the case of impersonal stance bundles, they predict events that are outside the speaker's sphere of influence (e.g., *is going to be* in a mathematical proof or geology). Finally, the ability bundles are used to express that the speaker or other participants of discourse are able to perform some action. As might be expected, they are often used with a second person pronoun to identify students' skills or distribute tasks (*[I want you] to be able to, [you have] to come up with*) (Biber 2006, 142).

5.2. Lexical bundles as discourse organizing devices

Biber identifies two major groups within this category: *topic introduction/focus* and *topic elaboration/identification*. It should be emphasized at this point these roles are not mutually exclusive and some of the discourse organizing bundles also function as stance expressions (Biber 2006, 142).

5.2.1. Topic introduction/focus bundles

The function of these bundles is to overtly signalize that a new topic is being introduced. As they are often expressed as a desire to open a new topic, there is some overlap between this category and the desire stance bundles (*e.g., I want to talk about, what I want to do*). These bundles are used with a first- or second-person pronouns and the plural *we* often suggests participation on the part of the audience (*[we are] going to talk about, [we] want to talk about*). They are sometimes expressed conditionally or tentatively (*if we look at, if you look at*). Finally, the bundles starting a *wh*-question represent a direct way to engage the audience in a new topic (*what do you think[?]*) (Biber 2006, 143).

5.2.2. Topic elaboration/clarification bundles

The bundles in this category serve to clarify a topic or introduce an additional explanation (*[it] has to do with the, you know I mean*). The two common and important bundles, *on the other hand* and *as well as the* belong to this category, their function being direct comparison and contrast. As demonstrated in the previous section, they are both used in academic prose rather than conversation and correspondingly, Biber found them to be used significantly more frequently in textbooks than in classroom teaching (Biber 2006, 144).

5.2.3. Identification/focus bundles

This type of bundles is sometimes categorized under referential, as their function is to mark the phrase following the bundle as particularly important. For example, *for those of you who [came late]* focuses the attention on a part of students. The bundles are often used after a lengthy explanation to conclude the main issues (*and that's one of [the problems], and this is a [real simple way]*), but sometimes an introduction bundle precedes a detailed explanation, focusing on the main point first (*one of the things [they stress in parenting]*) (Biber 2006, 144).

5.3. Referential bundles

Referential bundles generally identify an entity or some of its attributes as the temporary focus or an important part of the discourse. Biber distinguishes four subcategories: (1) *identification/focus*, (2) *imprecision indicators*, (3) *specification of attributes* and (4) *time, place or text reference* (Biber 2006, 145). The identification and focus bundles, the first type, have already been dealt with in the previous section because they fulfill a discourse organizing role as well.

5.3.1. Imprecision bundles

The imprecision bundles either indicate that the reference is not exact (*or something like that*) or suggest that additional references could be provided, but they are not the focus of the current discourse (*and things like that*) (Biber 2006, 145). It might be expected that spoken discourse is the domain of imprecision bundles due to their informality. Moreover, because they are fixed, easy and vague phrases, they can be used as fillers instead of meaningless sounds (cf. *evaluations, uhh ...* vs. *evaluations and things like that*). Because they are used to express the speaker's attitude towards the relevance of the signified or the precision of the reference, they can also be considered epistemic stance markers. (Biber 2006, 140).

5.3.2. Bundles specifying attributes

The bundles of this subcategory can identify various attributes, such as quantity (*have a lot of, a little bit of*), size or form (*the size of the [community], the form of half-wheels*), or abstract characteristic (*the nature of the [problem]*). The bundle *a little bit* is often used in the more specialized function of topic introduction, apparently to downplay the extent of the discourse in order to make the matter at hand seem more palatable (*[I want to talk] a little bit about [process control]*). In contrast, the more abstract bundles often establish logical relationships within the discourse (*on the basis of, in terms of the*) (Biber 2006, 146).

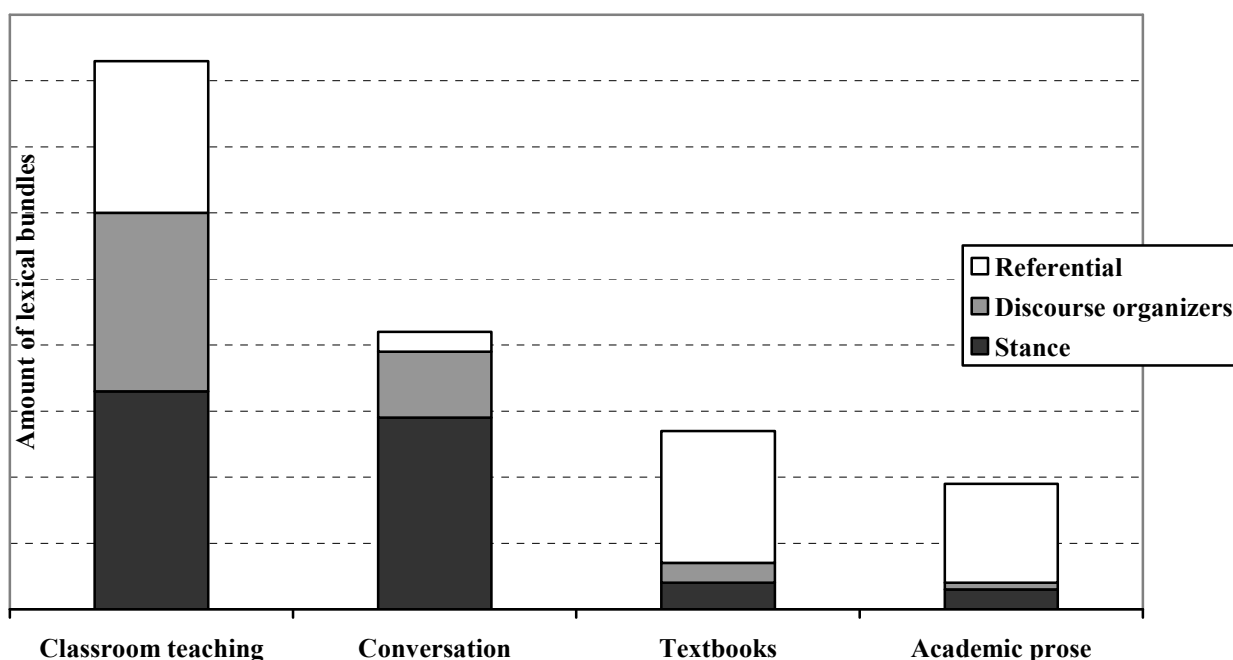
5.3.3. Bundles referring to time, place or text

Many of the the lexical bundles in the category have more than one function. For example, the bundle *the end of the* is used to refer to time (*the end of the [year]*), place (*the end of the [hall]*) or deictically to a part of the discourse (*the end of the [chapter]*). The deictic bundles are more common in the written registers, where they often point at a particular table or figure contained in the text (*as shown in figure 4.4*) (Biber 2006, 146).

5.4. Distribution of lexical bundles in different types of discourse

Even though Biber's study works with a different corpus (T2K-SWAL), his statistical data is useful not only as an illustration of how lexical bundles are used in different genres of spoken and written discourse, but also as a point of reference and comparison in the following sections. Biber's findings confirm the conclusion of other studies that classroom teaching tends to employ both "oral" and "literate" elements (Nasi and Basturkmen 2006, 290). The stance bundles are very common in conversation but rare in academic prose, while the opposite can be said about the referential bundles, and these two types are distributed almost equally in the classroom teaching (Biber 2006, 147). One of the intuitively plausible explanations might be that teaching relies on conversational elements because of the spoken form as well as intended accessibility and at the same time shares some the features of textbooks and academic prose (references and discourse organization) because the aim is fundamentally the same. Finally, research revealed that classroom teaching, in fact, employs more bundles of each category (stance, discourse organization and reference) than conversation, textbooks and academic prose, respectively (see figure 5.4.). The major aims and generally accepted quality criteria that are commonly applied on teaching, namely accessibility, clarity, good organization and contact with the audience, might serve as a tentative intuitive explanation of such abundance. This section concludes with a figure summarizing Biber's findings.

Figure 5.4.1. Proportional distribution of lexical bundles across functional categories according to Biber's study (Biber 2006, 147)



6. The cohesive role of lexical bundles

Lexical bundles can be also studied with regard to their role in textual cohesion, which is established by a number of grammatical and lexical means that form the surface links or connections within a text. The cohesive function of conjuncts, such as *however*, *moreover* or *nevertheless*, has been recognized for a long time, but some of the more recent studies also include multi-word prepositional phrases (*in addition*, *as a result of*) or even multi-word clause fragments with a verb (*which is to say* or *while it is true*) (Nesi and Basturkmen 2006, 284). As all of the multi-word examples above appear as recurrent expressions, mainly in academic prose, it becomes apparent that lexical bundles play a cohesive role that can be systematically studied. Many lexical bundles are used as “linking adverbials”, which typically express “intrasentential relations” (Nesi and Basturkmen, 2006, 284), and they can be divided into six major groups according to their function: (1) *enumeration/addition* is used to list items or add new items to the discourse; (2) *summation* concludes the preceding discourse; (3) *apposition* marks equivalence or inclusion; (4) *result/inference* indicates consequence; (5) *contrast/concession* expresses

comparison between two dissimilar objects; and (6) *transition* signals a change of topic or discontinuity with the preceding discourse.

The cohesive function of lexical bundles in academic lectures is discussed by Nesi and Basturkmen (2006) with the result that lexical bundles cover the whole range of above-mentioned cohesive functions: addition or apposition (*that's one of the, one of the things, and this is the*), summation (*at the end of the day*), concession and contrast (*while it is true, on the other hand*), result and inference (*as a result [off]*), and finally transition (*if you look at or going to talk about*) (Nesi and Basturkmen 2006, 284).

While the categorization based on the function of lexical bundles in the discourse and the one based on their cohesive role to a certain degree overlap, not entirely without confusion (Nesi and Basturkmen state, for instance, that the greatest potential for cohesive bundles is in the category of discourse organizers and referential bundles (Nesi and Basturkmen, 2006, 287)), both approaches will be considered in the discussion of lexical bundles in academic lectures.

7. Research of the lexical bundles in academic lectures

The following sections of this work present the data obtained from two sets of British and American lecture transcripts, analyze the discovered lexical bundles and discuss the findings in the context of previous studies. Before the data are presented, the following paragraphs specify the research scope and methods.

7.1. Research scope and methods

The present study investigates the use of lexical bundles in academic lectures recorded in two linguistic corpora, namely British Academic Spoken English (BASE) and Michigan Corpus of Academic Spoken English (MICASE). This paper does not focus on the distribution of lexical bundles across various fields and disciplines. Rather, it attempts to analyze various types of lexical bundles, their distribution and functions in academic lectures and discuss possible differences between British and American academic English in the area of lexical bundles with a high rate of occurrence. Only transcripts in Humanities and Arts and Social Sciences and Education were admitted, as these fields are generally more intelligible and as such more suitable

for demonstrating the functions of lexical bundles. Because the British lectures found in the BASE corpus are somewhat shorter than the American, the selection includes 14 lecture transcripts from BASE as compared to 11 transcripts from MICASE. To identify lexical bundles and discover the rate of their occurrence, the present work uses WordSmith Tools 5.0. It should be noted that some lectures found in these corpora include interactional content, such as short questions and answers, greetings and requests addressed to the lecturer to repeat a particular phrase or sentence. Because the scope of the present work is limited to academic monologue produced by a single speaker, highly interactive lectures were omitted and interactive contents were edited out from the rest. The selection from the British corpus amounts to approximately 103 000 words, similar to 107 000 words of the American selection.

It has been noted that most researchers define lexical bundles as expressions occurring more than 5 to 40 times per million words, the lower number being used when longer expressions are considered, whereas the higher limit represents a more conservative approach. In a corpus of approximately 100 000 words, this would mean that an expression has to occur at least once according to the lower standard – which is not a meaningful definition of a recurrent expression – or four times if a stricter limit is applied. It is obvious that a smaller corpus may significantly distort the results due to speaker idiosyncrasies and phrases limited to a particular subject, which is the reason why this study considers only expressions occurring at least 5 times in the selected transcripts and concentrates on the most common and most representative examples. With regard to speaker idiosyncrasies, each bundle has to occur at least in two lectures within the respective British and American sub-corpora, but some bundles present in only one transcript will be discussed nonetheless. Similar to research papers mentioned in the theoretical introduction, the focus will be on four-word bundles but both shorter and longer expression will be analyzed as well.

7.2. Research data: overview of the most frequent bundles

The following tables summarize 50 most frequently occurring lexical bundles in the respective sub-corpora of British and American academic lectures sorted by the number of occurrences. As can be seen, only four-word bundles found in two or more texts are included and the idiosyncratic bundles as well as the bundles of different length are briefly discussed later.

The fifth column marks the bundle as having a structure characteristic of conversation (C), academic texts (A) or possibly both (A/C), which is explained in more detail in Section 7.3. and which relies on Biber's classification as presented in the Longman Grammar of Spoken and Written English.

The sixth column identifies the bundle's discourse function using Biber's typology, marking each bundle as expressing stance (S), making a reference (R) or functioning as a discourse organizer (D) with the secondary or alternative category specified in the parentheses. This classification is elaborated on in Section 7.4.

Rather surprisingly, only 22% of the 50 most frequently occurring four-word bundles in the selected British and American academic lectures overlap. These 11 bundles are: *the end of the, of the nineteenth century, one of the things, on the other hand, if you look at, at the end of, on the other hand, and you can see, the rest of the, at the same time and in a in a*.

Fig. 7.2.1. The 50 most frequent lexical bundles in the British sub-corpus

Order	Lexical bundle	Frequency	Texts	Type	Function
1	THE END OF THE	21	12	A/C	R
2	OF THE NINETEENTH CENTURY	16	5	A	R
3	ONE OF THE THINGS	16	5	A/C	D(R)
4	THE WAY IN WHICH	13	4	A	R
5	AND ONE OF THE	13	6	A/C	D(R)
6	ON THE OTHER HAND	13	6	A	D
7	IS ONE OF THE	13	8	A/C	D(R)
8	WHAT I WANT TO	12	8	C	S(D)
9	IN OTHER WORDS THE	11	7	A	D
10	IN TERMS OF THE	10	4	A	R
11	THE THE THE THE	10	4	C	D(R)

12	IF YOU LOOK AT	10	6	C	D
13	AT THE END OF	10	7	A/C	R
14	ON THE ONE HAND	9	2	A	D
15	IN THE CONTEXT OF	9	4	A	R
16	POINT OF VIEW OF	9	4	A	R
17	THE POINT OF VIEW	9	4	A	R
18	A LOT OF THE	9	5	A/C	R
19	AT THE START OF	9	5	A	R
20	THE IDEA OF THE	9	5	A	R
21	THERE'S A KIND OF	9	5	C	S(R)
22	IN THE FAERIE QUEENE	8	2	A	R
23	OF THE FAERIE QUEENE	8	2	A	R
24	WANT TO LOOK AT	8	2	C	S(D)
25	DRAW YOUR ATTENTION TO	8	3	C	S(D)
26	FROM THE POINT OF	8	3	A	R
27	I'M NOT GOING TO	8	4	C	S(D)
28	AND YOU CAN SEE	8	5	C	S(D)
29	AS A KIND OF	8	5	A/C	S(R)
30	AND THIS IS THE	8	6	C	D(R)
31	CREATE A SENSE OF	7	2	C	R
32	I WOULD LIKE TO	7	2	C	S(D)
33	GOING TO BE THE	7	3	C	S(R)
34	IN THE EIGHTEENTH CENTURY	7	3	A	R
35	THE START OF THE	7	3	A	R
36	JUST TO REMIND YOU	7	4	C	S(D)
37	THE REST OF THE	7	5	A/C	R
38	THE WAYS IN WHICH	7	5	A	R
39	AT THE SAME TIME	7	6	A	D
40	IN A IN A	6	2	C	D(R)
41	ON THE OTHER SIDE	6	3	A/C	R
42	WAS GOING TO BE	6	3	C	R
43	WE SAW LAST TIME	6	3	C	D
44	YOU HAVE TO REMEMBER	6	3	C	S
45	A KIND OF A	6	4	A/C	S(R)
46	IN THE FIRST PLACE	6	4	A	D(S)
47	OF THE KIND OF	6	4	A/C	S(R)
48	SO I'M GOING TO	6	4	C	S(D)
49	THE ER YOU KNOW	6	4	C	S(D)
50	THE HISTORY OF THE	6	4	A	R

Fig. 7.2.2. The 50 most frequent lexical bundles in the American sub-corpus

Order	Word	Frequency	Texts	Type	Function
1	THE END OF THE	31	9	A/C	R
2	AT THE END OF	20	7	A/C	R
3	ON THE ONE HAND	16	4	A	D
4	AT THE SAME TIME	14	4	A	D
5	WHEN IT COMES TO	13	2	C	D
6	TO MAKE SURE THAT	12	5	A	S
7	OF COURSE YOU KNOW	11	2	C	S
8	YOU CAN SEE THE	11	3	C	S(D)
9	IF YOU LOOK AT	11	6	C	D
10	IN THE IN THE	11	6	C	D(R)
11	ON THE OTHER HAND	11	6	A	D
12	I WANT YOU TO	10	2	C	S(D)
13	AND YOU CAN SEE	10	4	C	S(D)
14	IN A IN A	10	5	C	D(R)
15	IF YOU THINK ABOUT	9	4	C	D
16	THAT THIS IS A	9	6	C	R(D)
17	TO DO WITH THE	9	6	C	D
18	HE WAS ABLE TO	8	2	C	R
19	IN THE NINETEENTH CENTURY	8	2	A	R
20	ONE OF THE THINGS	8	2	A/C	D(R)
21	TAKE A LOOK AT	8	2	C	D
22	BY THE END OF	8	4	A	R
23	THE REST OF THE	8	5	A/C	R
24	FOR THE MOST PART	7	2	A	S
25	IN THE COURSE OF	7	2	A	R
26	IT SEEMS TO ME	7	2	C	S
27	IN THE MIDDLE OF	7	3	A/C	R
28	OF THE THINGS THAT	7	3	A/C	D(R)
29	THE MIDDLE OF THE	7	3	A/C	R
30	FOR THE REST OF	7	6	A/C	R
31	END OF THE NINETEENTH	6	2	A	R
32	END OF THE SECOND	6	2	A	R
33	OF THE NINETEENTH CENTURY	6	2	A	R
34	YOU THINK ABOUT IT	6	2	C	D
35	IF YOU IF YOU	6	3	C	D(R)
36	TO BE ABLE TO	6	3	A	S(R)
37	HAVE TO DO WITH	6	4	C	D
38	I DON'T KNOW IF	6	4	C	S

39	IN THE UNITED STATES	6	4	A	R
40	THE BEGINNING OF THE	6	4	A/C	R
41	A LITTLE BIT ABOUT	6	5	C	R(D)
42	HAS TO DO WITH	6	6	C	D
43	AFTER WORLD WAR TWO	5	2	A	R
44	IN THE EIGHTEEN THIRTIES	5	2	A	R
45	AS YOU CAN SEE	5	3	C	S(D)
46	BY THE NAME OF	5	3	A	R
47	DURING WORLD WAR TWO	5	3	A	R
48	I THINK YOU CAN	5	3	C	S
49	IN THE TWENTIETH CENTURY	5	3	A	R
50	IT TURNS OUT THAT	5	3	C	S(D)

7.3. Idiosyncratic bundles

Because the bundles occurring in only one transcript are disregarded from this point onwards for the sake of objectivity, at least a few comments should be made about their nature. The selected transcripts revealed no idiosyncratic bundles in the strict sense of the term, i.e. pertaining to a single individual and possibly considered unusual or eccentric by other speakers. Rather, most of the unlisted bundles were excluded because their use is limited to a particular subject (e.g., *the NHS* or *the forum of Mussolini* in the British lectures and *of the second century* and *to the Roman people* in the American lectures). However, a few of the excluded lexical bundles are clearly not field-specific (*in terms of what* and *er er er er* in the British sub-corpus and *more likely to be* in the American). As none of the bundles that are considered idiosyncratic for the purposes of this study appears particularly eccentric, this discrepancy is probably caused by a relatively small size of the chosen sub-corpora and it seems reasonable to conclude these expression might appear as lexical bundles in a larger set of transcripts. On the other hand, it is possible that even though some expressions are not idiosyncratic in themselves, a high rate of their occurrence might be characteristic of only one or few speakers. Therefore, such expressions can be considered lexical bundles in the speech habits of particular language users, but their overall occurrence does not qualify them as lexical bundles in general English or even specific registers. From the statistical point of view, 9 out of the 50 most frequent bundles were excluded as idiosyncratic from the British sub-corpus and 13 from the American sub-corpus. Again, this discrepancy was probably caused by the difference in the number of selected lectures.

7.4. Shorter and longer bundles

As has been noted in the theoretical introduction, three-word bundles are extremely numerous and they can be usually treated as parts of longer expressions. For instance, a frequent bundle *one of the* is a part of a common bundle *one of the things* in American lectures. Moreover, three-word lexical bundles can often be treated as a single lexical unit (*in spite of, the United States, World War Two, etc.*). For these reasons, this study focuses on four-word bundles and longer, similar to previous studies on the subject (LG, 1000). Regarding the longer bundles, the results of previous studies have been confirmed and there are substantially fewer lexical bundles consisting of more than four words. Only eight non-idiosyncratic five- and six-word bundles have been found in the British sub-corpus: *the point of view of, and one of the things, from the point of view, from the point of view of, at the end of the, at the start of the, of the ways in which, this is one of the*. Nine non-idiosyncratic longer bundles have been found in the American sub-corpus: *at the end of the, by the end of the, if you think about it, one of the things that, the end of the nineteenth, the end of the second, end of the nineteenth century, in the eighteen thirties and, the end of nineteenth century*. Some of these expressions illustrate that longer bundles are often a product of chained shorter bundles, such as [₁ *from the* [₂ *point of*] ₁ *view of*] ₂. As a reasonable compromise between their linguistic relevance and their frequency of occurrence, only four-word bundles shall be discussed from this point onwards.

7.5. Structural categorization

First, it should be noted that it is impossible to categorize lexical bundles as “academic” or “conversational” solely on the basis of the structural classification as described in Sections 4.1., 4.2. and 4.3., because of the overlaps among the types. For instance, in terms of its structure, the bundle *as a result of* belongs to type 11 of conversational bundles and type 3 of academic bundles, but it is clear that the expression is very rare in common conversation. Moreover, this division can never be completely accurate, because it is reasonable to assume that even expressions such as *on the other hand* and *in spite of the*, which are typical of academic prose, would still appear, though maybe not frequently enough to qualify as bundles, in a sufficiently large corpus of non-academic spoken English. Therefore, the distinction relies on the typical,

rather than mutually exclusive structural forms of the bundles found in conversation, which tend to be verbal, and academic prose, which tend to be mostly nominal or prepositional (LG 995). However, as there is a large number of non-verbal lexical bundles that are still typical of conversation, for instance *that sort of thing* (LG, 1012), and bundles that can be safely assumed to occur in many registers of both spoken and written English, such as *a lot of the*, *the end of the* (LG, 1012, 1015), the category of mixed bundles has to be introduced. Even though Biber's exhaustive list was consulted when a bundle proved difficult to categorize, some bundles had to be classified with regard to their lexical and semantic elements. Therefore, without the evidence of a large-scale corpus study, the results are somewhat subjective and may only serve as a rough estimate. The findings are summarized in the table below.

Figure 7.5.1. Structural classification of lexical bundles

	British lectures	American lectures
Bundles typical of conversation	40%	36%
Bundles typical of academic prose	38%	44%
Mixed bundles	12%	18%

These results support earlier findings that classroom teaching relies on both “oral” and “literate” elements (Nesi and Basturkmen 2006, 290), but unlike the research by Biber (2006) and Nesi and Basturkmen (2006), the present work is limited to investigation of non-interactive content, which might be expected to have fewer features characteristic of conversation. Nonetheless, even in academic monologue, the lexical bundles typical of conversation and academic prose were found in almost equal proportion in both American and British lectures. This analysis is, however, limited to a basic structural distinction and the discourse functions of the lexical bundles have to be discussed in more detail.

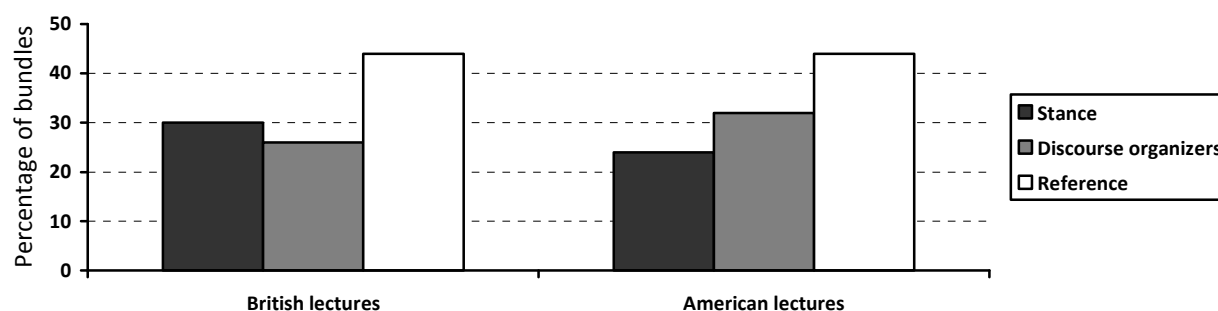
7.6. The discourse functions of the lexical bundles in the transcripts

As has been noted, the categories of discourse function may overlap. *I want you to [have a look at]*, for instance, can be interpreted as expressing a desire stance directed towards the audience or introducing a new topic. *And stuff like that* is a referential expression, but the reference is so vague that it might be plausibly considered an epistemic bundle expressing

uncertainty (Biber 2006, 143). Even though this ambiguity could be avoided by introducing a set of rules, it is only possible to discuss the discourse function of a particular occurrence of a bundle in a given context. Because the focus is on bundles with a high number of occurrences in several different texts, the context cannot be used to categorize the bundles according to their discourse functions. Therefore, some bundles had to be classified according to the presumed primary function, the cost of which is certain degree of unavoidable arbitrariness. For the sake of consistency, the bundles that express a desire or a decision were treated as stance bundles, even though they often serve to introduce a new topic (Biber 2006, 142). The expressions with *kind of* were taken as imprecision hedges rather than in the strict meaning of classification, as the former use is widely known to prevail in the spoken discourse. As in some other studies, (Biber 2006, 142) expressions such as *one of the things* were primarily treated as topic or focus articulation elements despite the reference they convey. Hesitation phenomena, such as *the the the the*, and fillers, such as *the er you know*, were treated as discourse organizers, as their primary function is to provide the speaker with additional time to formulate his or her thoughts. Moreover, hesitation marks and fillers might alert the audience to an imminent complex or unusual idea that takes time to formulate. While the present work acknowledges the problems in classification due to multiple functions of many lexical bundles and in particular cases discusses more than one function, the statistics below are based on the primary function for practical purposes.

Figure 7.6.1. Lexical bundles categorized according to their primary discourse function

	British lectures	American lectures
Stance	30%	24%
Discourse organizers	26%	32%
Referential	44%	44%



While these results largely support the findings of previous research, which revealed that classroom teaching employs a large number of bundles fulfilling the respective roles of stance and referential expressions as well as discourse organizers, Biber's recent study states that these three major types of bundles are distributed almost equally in classroom teaching (Biber 2006, 147 – see Figure 5.4.1.). On the other hand, the present data show that referential expressions form almost a half of all lexical bundles in both American and British transcripts. Two possible explanations might account for this discrepancy. First, research by Biber (2006) takes different styles of classroom teaching into consideration and as such includes also interactional content, which has been filtered out from the transcripts used in this study. Since it can be safely assumed that highly interactional parts of classroom teaching resemble general conversation, which uses far more bundles of stance than reference (Biber 2006, 147 – see Figure 5.4.1.), lower levels of interactivity might explain the higher proportion of referential bundles. Second, it is possible to speculate that if larger corpora had been used, some referential expressions would have had to yield to stance bundles or discourse organizers, as those tend to be more general and thus more likely to be used in a large set of lectures. To illustrate this point, the bundle *has to do with* can be used in almost any field and subject of study, whereas *during World War Two* is much more specialized and unlikely to appear among the 50 most frequently-occurring bundles in a large and diverse set of academic lectures. It is also clear that from the statistical standpoint, there is very little difference between the use of lexical bundles in American and British academic lectures (cf. vast differences between conversation and academic prose in Figure 5.4.1.). The discussion of particular bundles within each type might, therefore, prove more illuminating. It should be noted that in following examples, some rudimentary punctuation and capital letters were added to the transcripts when the division between sentences is obvious and unedited transcripts would be needlessly confusing.

7.6.1. Stance bundles

Regarding the epistemic bundles, British lectures seem to avoid direct personal epistemic stance as no such expressions have been found among the 50 most common lexical bundles. In comparison, American lecturers tend to express personal epistemic stance directly, such as *it*

seems to me, I don't know if and *I think you can*, the last being also an example of an ability bundle. Impersonal stance bundles are present in both British and American academic discourse, albeit of a different kind. Whereas British lectures employ a set of very similar nominal imprecision bundles that could be also categorized as referential expressions (*there's a kind of, as a kind of, a kind of a, of the kind of*), American impersonal bundles include a different qualifier, *for the most part*, and a verbal epistemic stance marker *it turns out that*.

Because the expressions with *kind of* are apparently referential, perhaps their classification as imprecision bundles requires justification by providing a few examples of their use:

...and the winner would go off to Rome to study in Italy and there we are so I'll just put up this as a kind of visual memento this is by Cogniet... (BASE, ahlct032_nopause.txt)

In this instance, the bundle serves to signalize figurative speech ("visual memento"). The reason for its use might lie in the fact that figurative speech is inherently imprecise because it is based on analogy and the speaker wanted to alert the audience before using such an expression. Another explanation is that the figurative expression required some forethought and the bundle *as a kind of* was used as a filler, thus functioning as a discourse organizer as well. In any case, it is clear that the role of the bundle is not referential despite the form, as the speaker did not mean a type or a category of a visual memento. Similar expressions, such as *there's a kind of*, are used in the same manner:

...er double double standard er we have to remember that the er the there's a kind of er difference in the way that er western Europe and its cultural... (BASE, ahlct002_nopause.txt)

Judging from the other fillers (*er*), at least part of the bundle (*kind of*) is used redundantly, again serving as a filler and imprecision marker.

It should be noted that in both British and American lectures, these expressions are used as hedges rather than boosters: with the exception of *it turns out that*, which is neutral to mildly affirmative, they add certain degree of uncertainty or qualification to the statement, particularly direct epistemic stance bundles in American lectures, such as *it seems to me* and *I don't know if*. It is reasonable to speculate that the reasons might be psychological: the teacher, while striving at authority, wants to create an atmosphere of shared exploration of knowledge and find common

ground with the students, who presumably feel much more uncertain in the subject than an expert. Moreover, the teacher might want to instill some epistemic modesty in the students by his example. Finally, the chosen fields of humanities and arts as well as social sciences are arguably more open to subjective interpretation than technical disciplines, which explains the epistemic bundles. To provide a few examples:

...have to engage with the real world but what then where could they go next. Now it seems to me that Twain's critical intelligence recognized the world as it was... (BASE, ahlct008_nopause.txt)

In the instance above, the hedge is used to alert the audience to the speculative nature of the discourse, as Mark Twain's perception of the world is, of course, not directly accessible. In the following example, the bundle *I don't know if* is used, beside its obvious meaning, to engage the audience:

...so if one considers the portraits and photographs for a moment and I don't know if you did consider them whilst reading the the novel... (BASE, ahlct013_nopause.txt)

Concerning the attitudinal bundles, British teachers tend to be more explicit about their wishes and intentions and many bundles expressing these categories are used: *what I want to*, *want to look at*, *I would like to*, *just to remind you*, *I'm not going to*, *so I'm going to*, *going to be the*. In the American transcripts, only one such bundle, *to make sure that*, appears among the 50 most common recurring expressions. It should be noted that many of the bundles, especially *what I want to*, could be categorized as discourse organizers, as they are used to introduce a new topic or focus on a particular angle. Because these bundles clearly express attitude and at the same function as discourse organizers, it must be admitted that in this case, the categorization is largely arbitrary. The following example illustrates how a bundle fulfills both these roles at once. Moreover, it also demonstrates how two four-word bundles *what I want to* and *want to look at* are combined into a five-word bundle *what I want to look at*:

...so what we've looked at in the last couple of lectures the process through which Spanish America became independent of European rule and what I want to look at is the process of nationalism broadly or to look at it another way I want to look at the question of what changed after independence... (BASE, ahlct007_nopause.txt)

In a similar use, the bundle *what I want to* is a part of an expression that announces topic elaboration and introduces a new idea:

...if you remember chapter five is the point when we've just got into the nineteenth century so what I want to suggest is that throughout Orlando Virginia Woolf parodies the style the literary style of the historical period she's talking about... (BASE, ahlct013_nopause.txt)

As for the requirements on the students, British teachers appear more assertive, using expressions *draw your attention to* and *you have to remember*, whereas the only bundle of this kind in American lectures was *I want you to*. American teachers, on the other hand, seem more ready to acknowledge the students' abilities by saying *of course you know*, *you can see the*, and *you can see*, the last bundle also occurring in British lectures. Beside the basic meaning of encouragement, these bundles are also used as discourse organizers when the speaker wants to summarize, clarify or elaborate on the subject and at the same time suggest that the following statements should be easily grasped in the context of previous discourse:

The example of a person in a chiefdom would be a chief, who, accepts contributions from everybody in a society and then redistributes it, according to the needs, of everyone in society. So redistribution, is another form, of exchange. So as you can see these two, reciprocity and redistribution, are not in keeping with the capitalist mode. (MICASE, LEL115JU090)

The bundle *the er you know*, which was found among the 50 most common in the British transcripts, was also included in the category of ability bundles because its surface meaning is to acknowledge the audience's understanding, but it often seems to merely fill the pause required for thinking:

...it's er was very very common to use non-professional actors again on the idea that somehow non-professional actors would get you past the kind of glamour system would get you past all the er you know all the kind of tricks and er theatricality and actorliness of actors... (BASE, ahlct015_nopause.txt)

Regarding the last example, it has been noted that because lexical bundles are theorized to be prefabricated, they are produced as a single unit and therefore, they are unlikely to be interrupted

by hesitation marks, fillers or pauses (Biber 2007, 283). On the other hand, it appears that a few entire four-word bundles, such as *the er you know* function as fillers. Similarly, some lexical bundles that can be considered hesitation marks, though not expressing stance, were found in British (*the the the the*) and American (*if you if you*) lectures, or both (*in a in a*). Such reduplicative bundles are, of course, typical of spoken discourse and as has been mentioned above, they can be considered discourse organizers despite the fact that they are often produced inadvertently, because they provide the speaker with additional time, help to avoid pauses and potentially prepare the audience for a complex idea or difficult expression. Such use can be demonstrated in the following example, in which the entire hesitation mark including the bundle *in a in a* is underlined:

...so here we have a visual image of Elizabeth not only in charge but almost in a in a slightly er deified position as though she's the representative... (BASE, ahlct011_nopause.txt)

The word *deified*, which follows the hesitation, is quite rare (specifically, it occurs only twice in the entire British sub-corpus of more than 100 000 words) and specific in meaning at the same time. Thus, it seems likely that the speaker was looking for the right word while filling the pause with the repetitive bundle *in a in a*, the hedge *slightly* and *er* sound.

The bundles that express confidence in the students' abilities, which are more typical of American teachers, are also used to emphasize a well-known fact and partly seem to function as fillers. In the following example, all expressions that directly engage the audience are underlined to illustrate how the bundle *of course you know* fits in the discourse that employs several different engagement features:

...okay? but here you know, why? Because Germany did not, use, positive incentives you know uh try to work with Britain constructively but rather, induced fear. So they induced, Britain ultimately to resist, and to engage in behavior, that ultimately you know from from the German standpoint of course you know, it's definitely not gonna be acceptable. So, there's a tendency to switch from one undesirable behavior towards another. In the economic realm and probably more recently, we see this a lot in um, in negotiations... (MICASE, LES495JU063)

To conclude the discussion of stance bundles, American and British lectures use stance expression to a similar extent and in similar ways. Nonetheless, it appears that there is a tendency

to avoid direct personal epistemic evaluations in British academic discourse, at least as far as the four-word bundles are concerned, and employ impersonal nominal imprecision bundles instead. On the other hand, British teachers are more direct in formulating the requirements on the students as well as explicitly stating their intentions and plans. American teachers have a greater tendency to encourage the students by expressing confidence in their abilities and knowledge.

7.6.2. Discourse organizers

The expressions serving as discourse organizers are quite similar in British and American academic lectures. Apart from the bundles that introduce a new topic or a point of view that have been already discussed in the previous section, British teachers use expressions such as *if you look at* and *we saw last time* to start a new topic, return to something that has been discussed before or set focus on a particular aspect of the discussed subject. American academic discourse employs a wider range of bundles in this role: *when it comes to*, *if you look at*, *if you think about*, *take a look at*, *if you think about*, *you think about it* and *if you if you*. The last one might seem uncertain, but it was included in this category because it is followed by *look at* in half of all occurrences:

...um, we tend if you if you look at infant mortality and and perinatal mortality statistics, they serve as kind of a gloss for how advanced a civilization is it's really interesting if you if you uh, if you look at those sorts of statistics... (MICASE, LEL115SU005)

On the other hand, the same or similar conditional expressions are sometimes used as an attempt to engage the audience. In the following example, all the audience-engagement features are underlined to illustrate how the bundle fits in the overall mode of presentation:

...well they placed ceilings on the number of missiles, but ins- instead of just producing missiles and single warhead missiles, you know they just put more warheads on one single missile. so you have a ceiling based, you may have one thousand missiles instead of two- two thousand missiles, but then if you if you place you know twelve warheads or ten warheads on on one missile instead of one warhead, I mean, in in in the end, I mean you have more warheads or more firepower available than you had before. (MICASE, LES495JU063)

Rather unsurprisingly, both American and British English use the pair *on the one hand* and *on the other hand* and the bundle *at the same time* for explicit comparison and contrast. However, whereas in the British transcripts, *on the other hand* is more numerous than *on the one hand*, the situation is reversed in the American sub-corpus (see Figures 8.1. and 8.2.), which means that comparison and contrast is expressed by just one disjunct from the pair or, alternatively, it may be paired with a different disjunct or conjunction. In the following example, the potential disjunct or conjunction is provided in brackets:

...so on the one hand I'm not saying you must read this. [but / on the other hand / nonetheless] I'm saying these are good starting points for reading and I expect you to explore and obviously essays which show that people have really explored explored the library explored the CD-ROMs in the library and followed up on things they've found through. (BASE, ahlct015_nopause.txt)

Beside these expressions, elaboration and clarification bundles *in other words* *the* and *in the first place* were found in the British transcripts, the latter often being used for emphasis, meaning “primarily” rather than “chronologically at the beginning”:

... and so he goes in and finds this monster and attacks this monster and of course it is his own mistake. One of the reasons he gets so tangled up with Errorr in Canto one in the first place is that he doesn't have the wisdom to know that... (BASE, ahlct010_nopause.txt)

American teachers appear to prefer bundles such as *have to do with*, *has to do with*, *to do with* *the* to explicitly address a link between two or more issues or offer an explanation, as in the following example:

But there's another way that Courbet's art can be described as real also, and that has to do with his painting technique and that's why I've brought along the Burial at Ornans from eighteen forty-nine on the left and a close-up of it on the right. (MICASE, LEL320JU143)

Regarding identification and focus, American and British teachers again employ very similar bundles, usually centered around demonstrative pronouns, the noun “thing” or numeral “one”: *is one of the*, *and one of the*, *and this is the* (British lectures), *of the things that* (American lectures), *one of the things* (both). The example below, taken from a British transcript, illustrates how the

last bundle is used to identify one of the main points that become the focus of subsequent discourse:

...she is doing this kind of satire this parody this critique and one of the things she parodies is precisely our tendency to frame historical periods. So at the end of chapter four she writes and I've given you that as the eighth extract...(BASE, ahlet013_nopause.txt)

To summarize the discussion of discourse organizers, it appears that there are no major differences in their use in the American and British academic lectures beside the preference for different phrases. For instance, the phrase “to have to do with” is often used in various forms in the American academic English but very little in British and the opposite is true for “in other words”. Nonetheless, both occur in the other sub-corpus as well. It should be mentioned that if all the expressions such as *what I want to* were included in this category instead of the category of stance bundles, the discourse organizers would form more than one third of all lexical bundles in both British and American sub-corpora. Even without including those bundles, the frequent occurrence of discourse organizers is one of the “oral” features of academic lectures, as they are characteristic of conversation rather than academic prose and textbooks (see Figure 5.4.).

7.6.3. Referential expressions

In case of referential expressions, the results are more difficult to summarize, because they include approximately a half of all bundles in both American and British lectures. As referential expressions, they tend to depend more on the field and topic than stance bundles and discourse organizers. Nonetheless, the same bundle, *the end of the*, occurs with the highest frequency in both sub-corpora. This might be easily explained by its multi-functionality, as it can refer to time, place or discourse-deixis (Biber 2006, 146). However, in the American transcripts, it referred to time in all 28 occasions, even though the temporal reference is sometimes oblique:

Okay. So to summarize by the end of the Civil War, the patterns of women's work, were set really up until...(MICASE, LEL105SU113)

In the British lectures, there is a minority of more abstract uses of the bundle, in the following example referring to the final part of an intangible object:

...here we will have we have evidence of a ruler called Tincommius. The similarity with the name of Commius provided an obvious dynastic link with all the coins themselves of this ruler we get as far as Tincom and that actually seemed to be very good in so far as it took us one stage further but it didn't supply the end of the name. In nineteen-ninety-six in Hampshire was found a hoard of coins produced by Tincommius but unfortunately or fortunately rather the name was complete and we now know this character is Tincommarus which basically translates as big fish an important man obviously so that is one of the problems... (BASE, ahlct006_nopause.txt)

Beside *the end of the*, both British and American lectures include many other bundles specifying attributes such as *the way in which*, *in the context of* (British lectures) or *in the course of*, *the beginning of the* (American lectures). Some of these bundles “establish logical relations within the text” (Biber 2006, 144), for instance *in terms of the*, which could also be categorized as a discourse organizer because it introduces a new aspect or sets focus on a particular angle:

...but er BBC were showing Chicago Hope er so they can be er they were seen as sexy and post-modern medical dramas yeah sexy and post-modern and of course there's that playful use of the word addicted yeah so a-, at once we get the glamorous address of those programmes in terms of star appeal and in terms of the seductive visualization of medical procedure now with reference to Cardiac Arrest. I want to argue that this series is is transitional... (BASE, ahlct014_nopause.txt)

This bundle is used in the same way in the American lectures, but not frequently enough to appear among the 50 most common bundles. In the following example, it specifies the point of view from which the issue will be approached:

...so risk is a very very rich kind of concept uh and i thought that it was something that we might want to take a look at not just, for today in terms of epidemiology, but in terms of the readings that we have been doing and uh and will be doing. For most epidemiologists risk exists as a statistical construct. (MICASE, LEL115SU005)

In both British and American lectures, referential bundles are also used for discourse-deixis, provided that the entire course can be considered a single large discourse divided into individual units, as the speakers often refer to previous or upcoming lectures:

...the real nation was something that we'll look at for the remainder of the year that's going to be the project I think for the rest of the year. Now I want to stop my lecture there but I don't want you to go away yet... (BASE, ahlct007_nopause.txt)

While this reference is on the surface temporal rather than deictic, the word *year* could be replaced by *course*, which consists of the academic lectures. In some cases, the deictic reference is quite clear and explicit:

And we're going to keep hearing this story we've heard it with Courbet we hearing it with Manet we're gonna hear it more and more for the rest of the class. The artist who shocks people. (MICASE, LEL320JU143)

A vast majority of the referential bundles falls into the category of “academic” bundles in terms of structure, as they tend to be nominal and often end with a preposition, such as *in the course of, the middle of the, in the nineteenth century, etc.*

Most of the differences between American and British lectures in terms of the referential bundles seems to stem from the differences in field and subject rather than a genuinely different use of language and as has already been noted, a larger corpus might eliminate many of the referential bundles from the list of the 50 most frequent expressions, such as *of the nineteenth century, in the Fairie Queene, of the Fairie Queene, in the eighteenth century* and *the history of the* in the British lectures and *in the nineteenth century, end of the second, in the eighteen thirties* in the American.

7.7. Cohesive role of lexical bundles in academic lectures

The research done by Nesi and Basturkmen (2006) showed that lexical bundles play a significant role in textual cohesion and a majority of the most frequent four-word lexical bundles in academic teaching signal cohesive relationships (Nesi and Basturkmen 2006, 293). The

present study supports these findings, as both the British and the American sub-corpus include frequent lexical bundles that function as cohesive devices. Moreover, each of the main intrasentential relation types, which are enumerated and explained in Section 6, is represented by a number of recurrent expressions in both sets of transcripts.

Enumeration and addition is often expressed by bundles such as *one of the things*, *and one of the* or *is one of the*:

...so here you see prejudice in its pure untainted form. er one of the things i want you to d-, to just er draw your attention to er here is that for Knox to have a queen on the throne and remember he's talking about Catholic queens...(BASE, ahlct011_nopause.txt)

It should be noted that in this example, the bundle *one of the things* also serves as a transition to a new topic or a new angle.

One of the typical expression of apposition is *in other words the* and *and this is the* in the British lectures. In the American sub-corpus, bundles used to express appositive relationship include *you can see the* and *and you can see*:

Um, nonetheless he takes, a new name, as a result of his victories he's beginning to sort of strike out on his own, and he takes the name victorious general, that's what imperator means so you can now see when you meet him on the street you could say oh how are you? Uh you are the Victorious General Son of a God Caesar, uh how very nice for you. Um, and you can see how Roman names can be used to carry a message. Um, and the way you can alter your name is partially a political message. (MICASE, LEL215SU150)

Summation is expressed in bundles such as *one of the things*, which in different contexts expresses apposition or addition:

...the effect of all this in terms of Rome and the effect of what the intention is one of the things we talked about very beginning today... (BASE, ahlct004_nopause.txt)

Result and inference is sometimes marked explicitly, for instance *so I'm going to* in the British lectures:

Er you need a motive to write and you have to choose the basest motive 'cause the romantics have taken up the noblest motives yes and the basest of motive is when I'm short of money and I think if I write a novel of say hundred-and-fifty pages I get a hundred pounds a page yeah fifteen-thousand pounds royalty so I'm going to go to the library and make a hundred pounds today... (BASE, ahlct001_nopause.txt)

Several bundles expressing concession or contrast can be found in both sets of transcripts, most notably *on the other hand* and *at the same time*. The latter retains the temporal meaning, marking concurrent phenomena or events, but it is often used to express contrast:

So this constant sort of tension between revealing, using the photograph to sort of reveal the specificity of identity, at the same time sort of denying you access, to that actual individual... so it becomes this just sort of implied presence. (MICASE, LEL320JU147)

Several transitional bundles can be found in both transcripts, such as *take a look at*, *if you look at*, *what I want to* or *if you think about*. It should be noted that rather than to mark radical departure from the established topic, these bundles are often used to introduce a new angle or provide illustrative examples, such as in the following example:

...and very often those snippets of information have an ulterior motive. If you look at the way that the source deals with Suetonius Paulinus and s-, Classicianus it's perfectly clear that Tacitus takes the side of the military figure... (BASE, ahlct005_nopause.txt)

These expressions have already been discussed in the section dedicated to lexical bundles as discourse organizers, which suggests that the analysis of cohesive and the discourse-organizing function of might partly overlap. Nesi and Basturkmen discuss cohesive bundles in terms of surface links within the discourse, or in other words, elements that express “intrasentential relations” and can be either considered linking adverbials or fulfill the same role (Nesi and Basturkmen 2006, 284). Similarly, Biber defines discourse organizers as expressions that “reflect relationships between prior and coming discourse” (Biber 2006, 135). Moreover, the elaboration of these concepts reveals more connections: Biber’s topic introduction and focus bundles are mostly additive, enumerative, transitional from the perspective of discourse cohesion (*if you look at*, *and this is the*, *what I want to*); topic elaboration and clarification bundles often play the

cohesive roles of apposition or summation (*and this is the*), result or inference (*so I'm going to*), concession or contrast (*on the other hand, at the same time*); identification and focus bundles are again appositive or enumerative (*and this is the, one of the things*). On the other hand, some bundles that have been categorized as primarily expressing reference (*at the end of*) or stance (*so I'm going to*) can function as cohesive devices. Conversely, some bundles that Biber ranks among discourse organizers (*has to do with*) (Biber 2006, 143) cannot be replaced by a suitable linking adverbial, which is a transparent criterion for assigning cohesive roles to multi-word phrases. (Nesi and Basturkmen, 293).

On the whole, the analysis of the data from the perspective of discourse cohesion yields very similar results as the discussion of lexical bundles as discourse organizers. It has been demonstrated that academic lectures rely on the cohesive devices because of the communicative aims of clarity in terms of both contents and structure. The data show that the differences between British and American academic lectures regarding the use of lexical bundles as cohesive devices are not systematic and stem from different preferences for particular phrases or the relatively small size of the chosen sub-corpora.

8. Conclusions

The aim of the present work was twofold: first, to provide a general overview on the subject of lexical bundles as a relatively recent area of study in corpus linguistics, explain the relevance of the concept and outline a structural and functional typology. Second, the study attempted to analyze the occurrence, use, structure and functions of the lexical bundles found in two sets of academic lectures, one British and one American.

Overall, the study supports the findings of previous research in several key points. It has been demonstrated that lexical bundles are units that usually “cut across grammatical structures” (Nesi and Basturkmen 2006, 286), which might explain why they had eluded the attention of linguists until recently. Regarding the use of lexical bundles in academic lectures as a specific genre of discourse, the study confirmed that academic lectures employ both bundles typical of conversation and those typical of academic prose in terms of their structure. As for their functions, the data show that the almost even proportional distribution of stance expressions, discourse organizers and referential bundles identify academic lectures as a specific genre. The

present work confirms the importance of this area of study, as the communicative aims of academic lectures combining accessibility and clarity with challenging density of new information reflect the balance between the stance bundles, references and discourse organizers.

The specific nature of non-interactive academic lectures proved more decisive than cultural and linguistic differences between American and British teachers, because while the lists of 50 most common bundles overlap in only one fifth, the proportional distribution of the structural and functional types is almost identical. The differences that could be considered systematic occur mostly in the area of stance bundles, but it must be conceded that significantly larger corpora would be needed to confirm the preliminary results showing differences in the use of hedging bundles, encouraging comments and directives.

9. Summary of the work in Slovak (Zhrnutie práce v slovenskom jazyku)

Práca skúma opakujúce sa slovné spojenia v akademických prednáškach britskej a americkej angličtiny. Je rozdelená do dvoch hlavných častí: v prvej sa podáva stručný prehľad o súčasných poznatkoch v tejto oblasti, ktorá spadá do oboru korpusovej lingvistiky, zatiaľ čo druhá časť predstavuje samostatné bádanie s využitím dvoch menších súborov prednášok (tj. približne 100 000 slov) britskej resp. americkej akademickej angličtiny, ktoré vznikli ako selekcia z korpusov BASE (British Academic Spoken English) a MICASE (Michigan Corpus of Academic Spoken English) v oblasti humanitných vied.

Prvá časť práce definuje a charakterizuje „lexical bundles“, čo by dalo doslova preložiť ako zhľuky lexikálnych jednotiek, je však o niečo transparentejšie hovoriť o opakujúcich sa slovných spojeniach obsahujúcich tri a viac slov. „Lexical bundles“ sú definované iba štatistickým výskytom bez ohľadu na slovný druh, štruktúrnú ucelenosť výrazu či postavenie vo vete z hľadiska významových celkov. Hoci nie je vytvorený jasný konsenzus, aký častý výskyt postačuje na to, aby sa výraz dal považovať za opakujúce sa slovné spojenie, väčšina bádateľov stanovuje limit v rozmedzí 10 až 40 výskytov na 1 000 000 slov v prípade kratších spojení (tri až štyri slová) a 5 až 10 v prípade dlhších spojení (päť a šesť slov). Medzi časté opakujúce sa slovné spojenia tak patrí napríklad štruktúrne neúplný výraz *what I want to* alebo *do you want me*, ktoré sú príznačné pre bežnú konverzáciu, alebo *as a result of* či *on the other hand*, charakteristické pre učebnice a iné akademické texty. Hoci, ako už bolo spomenuté, tieto

opakujúce sa slovné spojenia typicky nepredstavujú ucelené výrazy, na príklade *on the other hand* je vidieť, že menšia časť týchto výrazov je z hľadiska štruktúry a syntaktického využitia kompaktný celok. Práca ďalej charakterizuje „lexical bundles“ ako zväčša neidiomatické, hoci nie je ťažké nájsť výnimky ako *the big picture* či *the bottom line*. Dôležitým javom je tiež reťazenie kratších spojení, ako napríklad *what I want you* a *what you want to do*, ktoré sa dohromady vyskutujú ako šesťslovné spojenie *what I want you to do*.

Dôležitosť tohto smeru bádania v rámci korpusovej lingvistiky badať aj na tom, že v priemere až tretinu diskurzu, či už konverzácie alebo akademického textu, tvoria opakujúce sa výrazy, ktoré pritom vo väčšine prípadov nie sú štruktúrne úplné a netvoria významový celok, vďaka čomu sa ne napríklad pri výuke jazyka nekladie dôraz, ktorý by si podľa niektorých bádateľov zaslúžili.

Práca predkladá základnú typológiu opakujúcich sa slovných spojení, a to po stránke ich štruktúry ako aj funkcie vo vete. Na základe Biberovho výskumu opisuje 14 štruktúrnych typov, ktoré sú príznačné pre konverzáciu a 12 typov charakteristických pre písanú akademickú angličtinu. V záujme stručnosti postačí uviesť, že konverzácia sa do veľkej miery vyznačuje opakujúcimi sa spojeniami, v ktorých vystupuje slovesná fráza (*I want you, thank you very much*), zatiaľ čo akademické texty spravidla využívajú spojenia, ktorých základom je podstatné meno (*in the present chapter*) a často končia predložkou (*as a result of*). Treba však poznamenať, že toto delenie nie je úplné a vyskytujú sa napríklad opakujúce sa výrazy, ktoré štruktúrne pripomínajú skôr väčšinu „akademických“ spojení, avšak vyskytujú sa v temer všetkých registroch jazyka (*a lot of the*) alebo sú dokonca skôr príznačné pre konverzáciu (*for a couple of*).

Rozdelenie opakujúcich sa slovných spojení podľa ich funkcie sa taktiež opiera o Biberov výskum, ktorý rozlišuje výrazy postoja (*stance*), výrazy napomáhajúce organizácii prejavu (*discourse organizers*), ktoré vyjadrujú vzťahy medzi predošlými a nasledovnými časťami prejavu alebo textu, a referečné výrazy.

Výrazy postoja je možné ďalej deliť na epistemické, totiž vyjadrujúce postoj k pravdivosti či pravdepodobnosti výroku (*I don't know if*); výrazy osobného postoja (*I don't want to*); direktívne postoje (*you have to do*); výrazy, ktoré anticipujú budúci priebeh (*is going to be*) či touto formou vyjadrujú zámer (*what we're going to*); a na záver výrazy týkajúce sa možností a schopností participantov (*[I want you] to be able to*).

Výrazy napomáhajúce organizácii prejavu tiež zahŕňajú niekoľko podskupín: spojenia uvádzajúce tému či zameranie (*if we look at*), vysvetľujúce a rozvádžajúce výrazy (*has to do with, you know I mean*), a slovné spojenia vymedzenia či selekcie (*those of you who*).

Do referenčné slovných spojení sa zaraďujú „bundles“ slúžiace na vyjadrenie neistoty či nepresnosti (*or something like that*), ktoré sa dajú považovať aj za výraz postoja. Do kategórie referenčných spojení patria aj frázy pripisujúce vlastnosti či mieru (*a little bit of*). Veľkú podskupinu tvoria referencie časové, miestne, a deiktické (odvolávajúce sa na nejaký prvok diskurzu).

Význam tejto typológie odhaľujú štatistiky iných bádateľov, ktorí dokázali, že rôzne žánre, ako napríklad bežná konverzácia, učebnice, výuka a či výskumné rôzne žánre, sa vyznačujú zásadne odlišnou distribúciou týchto typov. Zatiaľ čo pre konverzáciu je charakteristické časté užívanie výrazov postoja, v článkoch a odborných pojednaniach prevládajú referenčné výrazy. Univerzitná výuka, ktorú skúma táto práca, je z tohto hľadiska výnimočná, keďže tri hlavné typy opakujúcich sa slovných spojení sú v tomto žánri zastúpené v takmer rovnakej miere a celkovým počtom prevyšujú iné formy ústneho či písomného prejavu.

V závere teoretickej časti sa práca venuje kohéznej úlohe opakujúcich sa slovných spojení, inými slovami, ako tieto výrazy slúžia ako základné spojivá textu či prejavu, ktoré vyjadrujú vzťahy medzi vetami alebo väčšími celkami diskurzu. Opäť sa rozlišuje niekoľko druhov kohéznych výrazov: enumeračné (pridávajúce nové prvky), zhrňujúce, apozičné (vyjadrujúce ekvivalenciu), kontrastívne a prípustkové, výrazy výsledku a následku a na záver slovné spojenia vyjadrujúce prechod k inej téme či zameraniu. Aspekt kohézie je pre bádateľov zaujímavý v tom, že veľká časť najpoužívanejších opakujúcich sa spojení zohráva úlohu v kohézii textu aspoň v niektorých kontextoch, a majú tak význam pre učiteľov aj študentov jazyka.

Výskumná časť práce začína náčrtom a zdôvodnením zvoleného postupu, najmä v ohľade zamerania na štvorslovné slovné spojenia a prednášky v oblasti humanitných vied. Hoci práca stručne charakterizuje aj kratšie a dlhšie slovné spojenia, podobne ako väčšina výskumných prác kladie dôraz na štvorslovné ako rozumný kompromis medzi dostatočným výskytom a relevantnosťou výrazu ako opakujúceho sa spojenia. Výber humanitných vied vysvetľuje ich väčšia miera zrozumiteľnosti a názornosti (aspon z pohľadu priemerného vzdelaného laika), keďže funkcie jednotlivých spojení sa ilustrovali na príkladoch.

Po vyradení idiosynkratických spojení (pre účely tejto práce takých, ktoré sa vyskytli len u jedného prednášajúceho) práca prezentuje 50 najpoužívanejších „bundles“ v britských a amerických prednáškach s frekvenciou výskytu, základným štruktúrnym rozlíšením na konverzačný či akademický typ a primárnou funkciou v diskurze.

Rozlíšenie spojení na príznačné pre konverzáciu alebo písomný akademický prejav na základe ich štruktúry komplikuje skutočnosť, že niektoré štruktúrne typy sa vyskytujú v oboch žánroch, vďaka čomu sa pri tomto delení muselo prihliadnuť aj na iné atribúty, ako napríklad stupeň formálnosti či konkrétny obsah spojenia. Bez rozsiahleho bádania v korpusoch bežnej hovorenej angličtiny a akademických textov má teda toto rozlíšenie skôr orientačnú podobu a bolo nutné zaviesť kategóriu zmiešaných spojení, ktoré sa vyskytujú v oboch žánroch prípadne nie sú jednoznačne zaraditeľné. Napriek tomu sa akademické prednášky preukázateľne využívajú spojenia príznačné pre konverzáciu ako aj písomný odborný resp. učebnicový prejav, a to dokonca v takmer rovnakej miere, čo platí pre britských ako aj amerických prednášajúcich.

Rozdelenie spojení do funkčných typov tiež nie je jednoznačné z dvoch hlavných dôvodov. Prvým je nejednoznačnosť v samotnej typológii podľa Bibera, ktorý sám pripúšťa, že niektoré výrazy plnia viacero úloh zároveň. Napríklad častý výraz *what I want to* je formálne výrazom priania či intencie prednášajúceho, typicky však uvádza novú tému či uhol pohľadu. Druhým dôvodom je to, že diskurzna funkcia výrazu je záležitosť konkrétnej inštancie, jedného výskytu daného spojenia. V prípade tejto práce sa však skúmajú práve spojenia s desiatkami rôznych výrazov v niekoľkých nezávislých prednáškach. Napriek nejednoznačnosti a istej miery arbitárnosti bolo 50 najčastejšie sa opakujúcich slovných spojení v britských aj amerických prednáškach roztriedených podľa primárnej funkcie. Potvrdil sa predpoklad, že hovorená akademická angličtina v hojnej miere využívajú všetky tri hlavné funkčné typy slovných spojení, avšak na rozdiel od výsledkov, ktoré udáva Biber, v prednáškach prevládali referenčné výrazy. Tento rozdiel sa dá vysvetliť tým, že Biber analyzuje univerzitnú výuku ako takú („classroom teaching“) a zámerne do výskumu zahrnul prednášky a semináre s rôznou mierou interaktivity, zatiaľ čo predmetom skúmania tejto práce sú len neinteraktívne prednášky. Dá sa pritom predpokladať, že čím vyššia je miera interaktivity, tým viac diskurz pripomína konverzáciu, kde prevládajú výrazy postoja. Z hľadiska týchto štatistík sa britská a americká akademická angličtina odlišujú len málo, z čoho vyplýva, že špecifickosť žánru akademickej prednášky je dôležitejší faktor než kultúrne a jazykové rozdiely.

Práca analyzuje konkrétne slovné spojenia rôznych typov a na príkladoch demonštruje ich využitie v prednáškach. V oblasti spojení vyjadrujúcich postoj sa britskí prednášajúci javia asertívnejšie, pokiaľ ide o formulovanie vlastných plánov a požiadaviek na študentov (*what I want to, draw your attention to*), na druhej strane sa zdržiavajú priamych epistemických hodnotení typu *it seems to me* a *I don't know if*, ktoré sa hojne vyskytujú v amerických prednáškach. Britskí prednášajúci majú tiež sklony využívať výrazy *s kind of*, ktoré signalizujú nepresnosť, vágnosť či neistotu (v anglickej terminológii sa nazývajú „hedges“), avšak občas len vyplňajú priestor, ktorý prednášajúci potrebuje na rozmyslenie.

V oblasti „discourse organizers“ alebo výrazov napomáhajúcich organizácii diskurzu sa britská a americká akademická angličtina líšia len málo. Na zmenu témy či zamerania využívajú výrazy ako *if you look at a one of the things* a na vyjadrenie porovnania či kontrastu využívajú okrem iných spojení aj známy idiomatický pár *on the one hand – on the other hand*. Do tejto kategórie patria aj prejavy váhania (*in a in a, the the the the*), keďže primárne poskytujú prednášajúcemu čas na rozmyslenie a potenciálne pripravujú poslucháčov na náročnejší slovný obrat, hoci pravdepodobne nie sú vždy vyslovované zámerne.

Oblasť referenčných výrazov je v takýchto malých korpusoch ťažké skúmať systematicky, keďže na rozdiel od výrazov postoja a organizačných prvkov diskurzu závisia na téme a predmete rozpravy. Okrem očakávateľných rozdielov, ako je napríklad častejší výskyt *in the United States* v amerických prednáškach a *the NHS* v britských (pozn. *National Health Service*, toto spojenie však bolo vyradené, keďže sa vyskytovalo len v jednej prednáške), sa dá konštatovať aj zaujímavá zhoda: najpoužívanejšie slovné spojenie v oboch korpusoch je *the end of the*, ktoré sa v britských aj akademických prednáškach v drvivej väčšine prípadov využíva ako časová referencia.

Posledným skúmaným aspektom bola kohézna úloha opakujúcich sa spojení. Tu sa opäť potvrdilo, že „lexical bundles“ sú významným kohéznym prvkom, keďže často vyjadrujú vzťahy medzi vetami a väčšími celkami a často preberajú funkciu adverbiálnych spájacích výrazov (linking adverbials), ako napríklad *and one of the things* zodpovedá *for example* a *in other words* *the* je jasne apozičným výrazom, ktorý pripravuje poslucháčov na parafrázu či obsirnejšie vysvetlenie. V oboch korpusoch sa navyše dajú nájsť príklady na všetkých šiest základných kohéznych funkciách. Treba poznamenať, že skúmanie „lexical bundles“ z hľadiska kohézie je často

podobné ich analýze z hľadiska organizácie diskurzu, keďže ide o príbuzné a vzájomne sa prelínajúce aspekty.

Výskumná časť práce priniesla niekoľko záverov. Ukázalo sa, že aj neinteraktívne akademické prednášky sú špecifickým žánrom hovoreného prejavu, ktorý spája jazykové prvky charakteristické pre ústny a písomný prejav, pre konverzáciu a odborné texty. Z hľadiska štruktúry využívaných opakujúcich sa spojení stoja prednášky presne na rozhraní medzi konverzáciou a akademickou prózou. Hojné zastúpenie všetkých troch základných funkčných druhov potvrdzuje, že špecifické komunikačné ciele a nároky akademických prednášok, totiž jasnosť, štrukturovanosť a zrozumiteľnosť spojené s veľkou hustotou nových informácií, sa prejavujú aj na úrovni najpoužívanějších opakujúcich sa spojení. Špecifickosť žánru sa pritom ukazuje ako podstatnejšia než jazykové či kultúrne rozdiely medzi britskými a americkými prednášajúcimi, ktoré v oblasti opakujúcich sa slovných spojení spočívajú skôr v rozdielnych preferenciách výberu fráz než systematických odlišnostiach. Na záver treba poznamenať, že také rozdiely môžu pri menších korpusoch založené skôr na individuálnych než jazykovo-kultúrnych preferenciách a konkluzívne výsledky v oblasti porovnania britskej a americkej akademickej angličtiny by vyžadovali niekoľkonásobne väčšie jazykové korpusy.

List of references and sources

References:

Douglas Biber and Federica Barbieri, “Lexical bundles in university spoken and written registers”, *English for Specific Purposes*, Vol. 26, Issue 3, pp. 283-304 (2007)

Ken Hyland, “As can be seen: Lexical bundles and disciplinary variation”, *English for Specific Purposes*, Vol. 27, Issue 1, pp. 4-21 (2008)

Douglas Biber, *University Language. A corpus-based study of spoken and written registers.* (Amsterdam and Philadelphia: John Benjamins Publishing Company, 2006)

Douglas Biber et al., *Longman Grammar of Spoken and Written English.* (Harlow: Longman, 1999)

Hilary Nesi and Helen Basturkmen, “Lexical bundles and discourse signalling in academic lectures”, *International Journal of Corpus Linguistics*, Vol. 11, Issue 3 (2006)

WordSmith Tools 5.0.0.179 by Mike Scott, Oxford University Press

Sources:

Michigan Corpus of Academic Spoken English (MICASE), <http://quod.lib.umich.edu/m/micase/>, last accessed on August 20, 2009

British Academic Spoken English (BASE), <http://www2.warwick.ac.uk/fac/soc/al/research/collect/base/>, last accessed on August 20, 2009

Lexical Bundles in Academic Lectures

(BA thesis abstract by Marek Leško)

The work attempts to briefly outline the current state of knowledge on the subject of lexical bundles and investigate their functions and use in two sets of academic lectures, British and American. For this purpose, it uses lecture transcripts from BASE (British Academic Spoken English) and MICASE (Michigan Corpus of Academic Spoken English) corpora.

In the theoretical part, lexical bundles are defined as recurrent expressions of three and more words that are identified by statistical means (at least 20 occurrences per million words, for instance) only with no regard for structural completeness or perceptual salience. The work presents general characteristics of lexical bundles as typically non-idiomatic, structurally incomplete units and sorts them into several structural categories typical of conversation or academic prose. Three main functions of lexical bundles are identified: stance expressions, discourse organizers and references. Finally, the study discusses lexical bundles in terms of cohesion, i.e. how these recurrent expressions are used as surface links within the discourse to signal mostly intrasentential relations that fulfill the role of linking adverbials.

In the practical (research) part, the thesis presents data obtained from two sets of lecture transcripts, British and American. The results include 50 most frequent lexical bundles in each sub-corpus together with their structural classification and primary discourse function. According to their structure, the bundles are categorized as “conversational”, “academic” or mixed. The data show that academic lectures are a specific genre of discourse, which employs both “oral” and “literate” elements. This conclusion supports the findings of previous research on the subject of classroom teaching. Regarding the functions of recurrent expressions, lectures include an abundance of bundles in each functional category, the majority being referential. Despite the prevalence of referential expressions, the categories are represented in a relatively balanced way, unlike in conversation, where stance expressions are prevalent, and academic prose, which use predominantly referential expressions.

The work presents individual examples of lexical bundles in each functional category and attempts to demonstrate differences between British and American academic English. In case of stance bundles, British lecturers seemed more assertive with regard to their audience but less likely to use direct epistemic evaluations. American teachers, on the other hand, seemed more

encouraging and several motivational expressions are used among the 50 most common lexical bundles. In the area of discourse organizers, British and American lectures employ the same or similar expressions to convey relationships between various parts of discourse. Referential bundles were more difficult to evaluate, as they tend to depend on the subject and the size of the corpora was not sufficient to discuss systematic differences.

Finally, the work demonstrates the cohesive role of lexical bundles on individual examples and confirms the results of previous research that recurrent expressions play a significant cohesive role in academic discourse.

It has been shown that academic lectures are a specific genre of discourse due to the communicative aims and demands combining clarity, good structure and a density of new information. The unique nature of this genre is a more important factor than linguistic or cultural differences between British and American lecturers.

Opakujúce sa slovné spojenia v akademických prednáškach (abstrakt k bakalárskej práci, Marek Leško)

Práca predstavuje stručný náčrt súčasných poznatkov o „lexical bundles“ (opakujúcich sa slovných spojení) a skúma ich využitie a funkcie v dvoch sadách akademických prednášok, jednej britskej a jednej americkej. Za týmto účelom pracuje s prepismi prednášok vybraných z korpusov BASE (British Academic Spoken English) a MICASE (Michigan Corpus of Academic Spoken English).

V teoretickej časti sa definujú „lexical bundles“ ako opakujúce sa výrazy troch a viacerých slov, ktoré sú určené iba štatistickým výskytom (napríklad dvadsiatimi výskytmi na milión slov) bez ohľadu na štruktúrne úplnosť či postavenie vo vete z hľadiska interpretácie. Práca predkladá všeobecnú charakterizáciu týchto spojení ako zväčša neidiomatických, štruktúrne neúplných jednotiek a triedi ich do niekoľkých štruktúrnych kategórií, ktoré sú typické pre konverzáciu alebo akademickú prózu. Tieto spojenia majú tri hlavné funkcie: vyjadrujú postoj, napomáhajú organizácii diskurzu alebo slúžia ako referenčné výrazy. Na záver teoretickej časti práca analyzuje „lexical bundles“ z hľadiska kohézie, teda spôsobu, akým prispievajú k prepojeniu nadvetných celkov podobne ako adverbiály s touto úlohou.

V praktickej, výskumnej časti práca predkladá dáta získané z dvoch sad prepisov britských a amerických prednášok. V tejto časti sa prezentuje 50 najčastejších opakujúcich sa spojení v každej sade spolu s ich zaradením na základe štruktúry ako aj primárnej diskurznej funkcie. Podľa štruktúry sa delia na „konverzačné“ a „akademické“, prípadne charakteristické pre oba registre. Výsledky ukazujú, že akademické prednášky sú špecifickým žánrom, ktorý využíva prvky typické pre ústny aj odborný písomný prejav. Tento záver podporuje predošlý výskum v tejto oblasti. Pokiaľ ide o funkcie opakujúcich sa spojení, prednášky zahŕňajú množstvo „bundles“ v každej funkčnej kategórii, pričom väčšinu tvoria referenčné. Napriek prevahe referenčných výrazov sú jednotlivé kategórie zastúpené relatívne vyvážené na rozdiel od konverzácie, kde prevládajú výrazy postoja, a akademickej prózy, ktorá využíva najmä referenčné spojenia.

Práca prezentuje konkrétne príklady opakujúcich sa spojení v každej funkčnej kategórii a pokúša sa na nich demonštrovať rozdiely medzi britskou a americkou akademickou angličtinou. V prípade výrazov postoja, britskí prednášajúci sú podľa výsledkov asertívnejší voči poslucháčom, ale zdržanliví v prípade priamych epistemických súdov. Na druhej strane pôsobia americkí prednášajúci viac motivačne a medzi 50 najčastejšími výrazmi sa objavuje niekoľko povzbudzujúcich fráz. V oblasti výrazov napomáhajúcich organizácii diskurzu využívajú britskí aj americkí prednášajúci podobné spojenia vyjadrujúce vzťahy medzi jednotlivými časťami diskurzu. Je ťažšie vyhodnotiť referenčné výrazy, keďže spravidla závisia na téme a veľkosť korpusov nestačila na analýzu systematických rozdielov.

Práca na záver na jednotlivých príkladoch demonštruje kohéznu úlohu opakujúcich sa spojení a potvrdzuje tvrdenia predošlého výskumu, že tieto výrazy zohrávajú v kohézii textu v prípade akademických prednášok významnú úlohu.

Ukázalo sa, že akademické prednášky sú špecifickým žánrom diskurzu vďaka komunikačným cieľom a nárokom, v ktorých sa kombinuje požiadavok jasnosti a dobrej štruktúry s hustotou nových informácií. Unikátna povaha tohto žánru je dôležitejší faktor než jazykové či kultúrne odlišnosti medzi britskými a americkými prednášajúcimi.